CENTRALIZED AND DISTRIBUTED RESOURCE ALLOCATION WITH APPLICATIONS TO SIGNAL PROCESSING IN COMMUNICATIONS

BY

ALBERTH ESTUARDO ALVARADO ORTIZ

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Industrial Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Associate Professor Angelia Nedich, Chair
Professor Jong-Shi Pang, Director of Research
Associate Professor Ramavarapu Sreenivas
Assistant Professor Gesualdo Scutari, SUNY at Buffalo

Abstract

Nowadays, wired and wireless networks are used everywhere and everyday. With the increasing popularity of multiuser communication systems, their optimal performance has become a crucial field of study during the last decades. A factor that greatly determines such performance is the optimal allocation of the resources available to the agents in the network. This dissertation provides a set of optimization techniques applicable to rigorously address and deeply analyze multiuser resource allocation problems in different areas, ranging from signal processing, to communications and networking. More specifically, this work focuses on the three main topics that we briefly describe next.

First, we study the maximum sum-utility achieved when a noncooperative approach is used to allocate the spectrum in a communication system adopting a dynamic spectrum management framework. In particular, we turn our attention to the case in which the users in the system are endowed with infinite power budgets. This asymptotic analysis, based on the linear complementarity problem theory, leads us characterize the behavior of the system's utility as the power budget is increased toward infinity, and thus draw interesting conclusions on the efficiency of the Nash equilibrium and the Braess-type paradox, among others.

Second, we propose a novel class of distributed algorithms for the optimization of nonconvex and nonseparable sum-utility functions subject to convex coupling constraints. Even though, we focus on utility functions of the Difference of Convex (DC) type, further generalizations are possible. Moreover, the obtained iterative schemes are provable convergent to stationary points of such optimization problems. Among the different applications of our Successive Convex Approximations based algorithms, we direct our attention to a novel resource allocation problem in the emerging field of

physical layer based security, and to the well-known MIMO (Multiple-Input-Multiple-Output) Cognitive Radio sum-rate maximization problem. For the former application, we develop a mathematically rigorous analysis of the nondifferentiable and nonconvex game (of the generalized type) proposed to optimally allocate the network resources in this context; and finally, we apply our algorithms to find relaxed equilibrium points of the mentioned game. For the second application, our theory provides, for the first time, a provable convergent algorithm.

The third major topic of this dissertation analyzes a multiuser maximization problem where the utility function has a particular structure, namely, it is the sum of continuous maximum functions, subject to private and coupling constraints. We follow two different approaches in order to design provable convergent algorithms to address this problem. These approaches are based on "simpler" reformulations of the nondifferentiable and nonconcave optimization problem of interest. A careful analysis relating such problems is also developed. The cited results pave the way to devise (possibly distributed) algorithms for different system designs in the context of physical layer based security, ranging from the secrecy sum-rate maximization to the Max-Min fairness problem. It is important to emphasize that, different from the simple networks models considered in the physical layer security literature, the system designs studied in this dissertation involve networks composed of multiple legitimate users and friendly jammers, and a single eavesdropper, where the main users communicate over multiple (either orthogonal or non-orthogonal) subchannels.

Finally, it is worth mentioning that most of the tools and results developed in this work are general enough to encompass applications in many fields different from those described above. This dissertation highlights how the introduction of optimization theory in different signal processing applications has motivated several significant developments in the former field, in particular in the area of multiuser distributed optimization. Future research directions are provided at the end of each chapter.

To God, for the strength, wisdom and patience to complete this research.
To my parents, for their unconditional love and support.

Acknowledgments

I would like to express the deepest appreciation to my advisor, Professor Jong-Shi Pang, for his excellent guidance, caring, patience and understanding throughout all these years. Of course, without his support this dissertation would not have been possible. The list of all the things I have learned from him is very long, among them, I would like to thank him for transmitting to me his passion and excitement towards research. Furthermore, besides the mathematics and engineering skills that I gained from each meeting with Professor Pang, he taught me, through experience, the techniques needed to enhance my research skills, and thus produce results of rigorous quality and great impact. In particular, he showed me that with patience, dedication, persistency and great effort nothing is impossible.

I would like to thank my committee members Professor Angelia Nedich, Professor Ramavarapu Sreenivas and Professor Gesualdo Scutari for their invaluable input and service. Particularly, I would like to thank Professors Nedich and Sreenivas for sharing with me their knowledge in the area of optimization through distinct graduate courses that I had the great opportunity to take with them. I was very fortunate to have such respectable and passionate teachers. Additionally, I would like to extend an especial gratitude to Professor Scutari for his great guidance, dedication, patience and the feedback provided during the development of this work. Without his expertise in the signal processing field, this dissertation would have never come to an end.

I would like to express my profound gratitude to my beloved mother, Delia de Alvarado, and to my father, Carlos Alvarado, for always being there, in spite of the distance that separated us, encouraging me with their best wishes. Their love and support was definitely my greatest motivation to accomplish this goal. I also want to acknowledge the unconditional help of my sister, Lisbeth Alvarado, and my brother, Carlos Alvarado; thank you very much for cheering me up. Unquestionably, I would not have had the strength to go through the different difficulties of graduate school without the guidance and long-distance company of all my family members.

I want to express thanks to all the staff of the Industrial and Enterprise Systems Engineering Department, for assisting me with the administrative tasks necessary for completing my degree. In particular, I want to thank the Office Administrator, Amy Summers, and the Graduate Programs Coordinator, Holly Kizer, for all their kind help provided during my studies at this institution.

Finally, I would also like to thank all of my former students during my years of instructor at Galileo University and Francisco Marroquin University in my home country Guatemala, for motivating me to pursue this degree. Your eagerness of knowledge motivated myself to go forth and acquire deeper preparation for the future generations to come.

Table of Contents

List of Symbols

Spaces

$\mathbb{R}^n$	the real n-dimensional space.
$\mathbb{R}$	the real line.
$\mathbb{R}^n_+$	the nonnegative orthant of $\mathbb{R}^n$.
$\mathbb{R}^{n\times m}$	the space of $n \times m$ real matrices.
$\mathbb{C}^{n\times m}$	the space of $n \times m$ complex matrices.

Scalars

x	the scalar x.
$\{x^\nu\}$	the sequence of scalars $x^1, x^2, x^3, \ldots$
x^+	the nonnegative part of a scalar, i.e. $x^+ \triangleq \max(0, x)$.
$\lfloor x \rfloor$	floor function i.e. the largest integer not greater than x.

Vectors

$\mathbf{x}$	the column vector $\mathbf{x} \in \mathbb{R}^n$.
$\mathbf{x}^T$	the transpose of vector $\mathbf{x}$.
$\{\mathbf{x}^\nu\}$	the sequence of vectors $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \ldots$
$\mathbf{x}^T\mathbf{y}$	the inner product of vectors in $\mathbb{R}^n$.
$\\|\mathbf{x}\\|_p$	the ℓ_p-norm of the vector $\mathbf{x} \in \mathbb{R}^n$.
$\\|\mathbf{x}\\|$	the Euclidean norm of the vector $\mathbf{x} \in \mathbb{R}^n$.
$\mathbf{x} \perp \mathbf{y}$	the vectors $\mathbf{x}$ and $\mathbf{y}$ are perpendicular.
$\mathbf{1}_n$	the n-dimensional vector of all ones.

$\mathbf{0}_n$ the n-dimensional zero vector.

Matrices

$\mathbf{A}$ the matrix with entries a_{ij} i.e. $\mathbf{A} \triangleq (a_{ij})$.

$\mathbf{A}^{-1}$ the inverse of matrix $\mathbf{A}$.

$\mathbf{A}^T$ the transpose of matrix $\mathbf{A}$.

$\mathbf{A}^H$ the Hermitian of matrix $\mathbf{A}$.

$\det(\mathbf{A})$ the determinant of matrix $\mathbf{A}$.

$\mathrm{tr}(\mathbf{A})$ the trace of matrix $\mathbf{A}$.

$\rho(\mathbf{A})$ the spectral radius of matrix $\mathbf{A}$.

$\|\mathbf{A}\|$ the Euclidean norm of matrix $\mathbf{A}$.

$\|\mathbf{A}\|_F$ the Frobenius norm of matrix $\mathbf{A}$.

$\mathbf{A}_{\bullet\alpha}$ the columns of matrix $\mathbf{A}$ indexed by the set α.

$\mathbf{A}_{\alpha\bullet}$ the rows of matrix $\mathbf{A}$ indexed by the set α.

$\mathbf{A}_{\alpha\beta}$ the submatrix of matrix $\mathbf{A}$ with rows and columns indexed by sets α and β.

$\mathbf{I}_n$ the identity matrix of order n.

$\mathrm{Diag}(\mathbf{a})$ the diagonal matrix with diagonal elements equal to the components of the vector $\mathbf{a}$.

$\mathbf{A} \succ \mathbf{0}$ the positive definite matrix $\mathbf{A}$.

$\mathbf{A} \succeq \mathbf{0}$ the positive semidefinite matrix $\mathbf{A}$.

Sets

$\in, \notin$ element membership, non-membership in a set.

$\subseteq, \subset$ set inclusion, proper set inclusion.

$\cup, \cap, \times$ union, intersection and Cartesian product of sets.

$\bigcup_{i=1}^n S_i$ the union of sets S_i.

$\prod_{i=1}^n S_i$ the Cartesian product of sets S_i.

$S_1 \setminus S_2$ the difference of sets S_1 and S_2.

$\overline{S}$ the complement of set S.

$|S|$ the cardinality of a finite set S.

$\emptyset$ the empty set.

$\underset{X}{\operatorname{argmax}} f(x)$ the set of constrained maximizers of the function f over X.

$\underset{X}{\operatorname{argmin}} f(x)$ the set of constrained minimizers of the function f over X.

$\mathcal{N}_S(\mathbf{x})$ the normal cone of the set S at $\mathbf{x} \in S$.

Functions

$\mathbf{F} : \mathcal{D} \to \mathcal{R}$ a mapping with domain $\mathcal{D}$ and range $\mathcal{R}$.

$f(\mathbf{x})$ a real valued function f evaluated at $\mathbf{x}$.

$f(\mathbf{x}; \mathbf{y})$ a real valued function f evaluated at $(\mathbf{x}, \mathbf{y})$ where $\mathbf{y}$ is treated as a constant parameter.

$\frac{\partial f(\mathbf{x})}{\partial x_i}$ the partial derivative of the real valued function f with respect to x_i evaluated at $\mathbf{x}$.

$\nabla f(\mathbf{x})$ the gradient of the real valued function f at $\mathbf{x}$.

$\nabla^2 f(\mathbf{x})$ the Hessian matrix of the real valued function f at $\mathbf{x}$.

$f'(\mathbf{x}; \mathbf{d})$ the directional derivative of the real valued function f along direction $\mathbf{d}$ evaluated at $\mathbf{x}$.

$\partial f(\mathbf{x})$ the subdifferential of the real valued function f at $\mathbf{x}$.

$\Pi_S(\mathbf{x})$ the Euclidean projection of $\mathbf{x}$ onto the set S.

VI Notation

$\mathrm{VI}(K, \mathbf{F})$ the variational inequality defined by the set K and the mapping $\mathbf{F}$.

$\mathrm{LCP}(\mathbf{q}, \mathbf{M})$ the linear complementarity problem defined by the vector $\mathbf{q}$ and the matrix $\mathbf{M}$.

$\mathbf{F}_K^{\mathrm{nat}}$ the natural map associated with the pair $(K, \mathbf{F})$.

Chapter 1

Introduction

1.1 Preliminaries

In the last years, optimization theory has been widely used in a wide variety of applications in signal processing, communications and networking. Among the different areas of optimization theory adopted by the aforementioned communities, we can mention convex and nonconvex optimization techniques [11], game theory [76, 80, 7], and more recently, the variational inequality (VI) approach [28] has also been applied to deal with equilibrium problems where classical methods may not be enough to address them; we refer the interested reader to [98] for a general overview of these topics. The mentioned optimization techniques have been extensively used to optimally allocate (limited) resources in communication systems. The analysis of different resource allocation problems in multiuser systems and the design of provable convergent algorithms attempting to solve them, conform the main interests of this dissertation. Despite the fact that some of the algorithms proposed in this work are centralized, our main interest is the design of distributed schemes, since there are several applications in the area of signal processing in communications where the centralized approaches may be infeasible or too demanding.

In this work, we are primarily concerned with two different approaches followed to optimally allocate resources in a multiuser system, namely:

i) the well-known Network Utility Maximization (NUM) framework (see, Subsection 1.1.1); and,

ii) game theoretical models (see, Subsection 1.1.2).

Needless to say, there is a vast literature related to these subjects. In the next two sections we briefly introduce them and point out relevant references to those readers interested in more details. These sections do not intend to be a comprehensive study of the aforementioned topics but rather a quick overview of them with the objective of setting the foundations necessary to easily understand the forthcoming chapters. After these topics have been introduced, we also set the context of the applications in signal processing to communications that are studied along this dissertation; this is covered in Subsection 1.1.3.

1.1.1 Network Utility Maximization

Since the introduction of the concept of NUM in the paper [49], it has been generalized to encompass different applications, ranging form Internet congestion control protocol to resource (such as power, bandwidth, rates, etc.) allocation problems in either wired or wireless communication networks.

Let us introduce the concept of NUM formally. For that, consider a multiuser system composed of I users. Each user $i = 1, \ldots, I$ makes decision on his n_i-dimensional real strategy vector $\mathbf{x}_i \in \mathbb{R}^{n_i}$ subject to some local constraints represented by the closed (nonempty) set $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$. Let the Cartesian structured set $\mathcal{X} \triangleq \prod_{i=1}^{I} \mathcal{X}_i$ denote the set of private constraints. Suppose also, that there are n_c coupling constraints of the form $\mathbf{h}(\mathbf{x}) \triangleq (h_j(\mathbf{x}))_{j=1}^{n_c} \leq \mathbf{0}$, where $h_j : \mathcal{X} \to \mathbb{R}$ for all $j = 1, \ldots, n_c$. Assume that every user obtains certain reward when choosing strategy $\mathbf{x}_i$, that is, each user $i = 1, \ldots, I$ has an utility function $U_i : \mathcal{X}_i \to \mathbb{R}$. Hence, a generalized version of NUM is the problem of finding the strategy profile of all the users in the system $\mathbf{x} \triangleq (\mathbf{x}_i)_{i=1}^{I}$ that maximizes the total utility of the system $U(\mathbf{x}) \triangleq \sum_{i=1}^{I} U_i(\mathbf{x}_i)$, the so-called sum-utility, subject to the set of private and coupling constraints, i.e. we aim to solve

$$\begin{array}{lll} \underset{\mathbf{x} \triangleq (\mathbf{x}_i)_{i=1}^{I}}{\text{maximize}} & U(\mathbf{x}) \triangleq \displaystyle\sum_{i=1}^{I} U_i(\mathbf{x}_i) & \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i \quad \forall\, i = 1, \ldots, I & \textit{(private constraints)} \\ & \mathbf{h}(\mathbf{x}) \leq \mathbf{0} & \textit{(coupling constraints).} \end{array} \tag{1.1}$$

For the sake of notational simplicity, let $\Xi \triangleq \{\mathbf{x} \in \mathcal{X} : \mathbf{h}(\mathbf{x}) \leq \mathbf{0}\}$ denote the feasible set of (1.1). It is worth remarking that many resource allocation problems can be formulated as a single constrained optimization problem of the general form introduced in (1.1) that seeks to maximize some suitably chosen utility function U.

There are also cases in which the utility of user i is assigned a particular weight. For example, in situations where the system wants to prioritize some users, then a weighted sum-utility function of the form $\widehat{U}(\mathbf{x}) \triangleq \sum_{i=1}^{I} w_i \, U_i(\mathbf{x}_i)$ can be used instead, where the scalar w_i represents the weight assigned to user i. For the sake of simplicity, in what follows we stay with the objective function in (1.1), that is, we let the weights $w_i = 1$ for all $i = 1, \ldots, I$.

So far, we have not impose any assumption with regard to the objective function or the constraints of the program (1.1). The simplest case occurs when such a resource allocation problem can be modeled by an utility function U that is concave on $\mathcal{X}$, the set of private constraints $\mathcal{X}$ is convex and there are no coupling constraints. A key characteristic of the sum-utility function in (1.1) is that it is separable in each users' variables, so is the set of constraints $\mathcal{X}$ (due to its Cartesian structure). These features along with the standard convexity assumptions mentioned above, significantly simplify the analysis of (1.1), moreover, it is possible to design distributed algorithms converging (under some mild conditions) to globally optimal solutions of the NUM problem (1.1), see, e.g., [12].

When the constraint set is coupled, distributed schemes can still be devised to solve (1.1). For example, under the same standard convexity assumptions of the objective and private constraints of (1.1), if the coupling constraints are in the separable form $\mathbf{h}(\mathbf{x}) \triangleq \sum_{i=1}^{I} \mathbf{h}_i(\mathbf{x}_i) \leq \mathbf{0}$ with each h_i being convex over $\mathcal{X}_i$, then dual or primal decomposition techniques can be invoked to design distributed algorithms. We refer the reader to [82, 83] for a detailed analysis regarding decomposition methods for NUM.

A different and more general case is that of coupled sum-utility functions U, that is, when the utility of the i-th user U_i is affected by the strategy vector of the rest of users in the system $\mathbf{x}_{-i} \triangleq (\mathbf{x}_j)_{j \neq i}$. This situation arises naturally in systems with cooperation or competition between their agents.

Under this assumption, the maximization problem (1.1) adopts the form

$$\underset{\mathbf{x}\in\Xi}{\text{maximize}} \quad U(\mathbf{x}) \triangleq \sum_{i=1}^{I} U_i(\mathbf{x}_i, \mathbf{x}_{-i}), \tag{1.2}$$

where we have both a coupled objective and coupling constraints. Assuming that the objective function U is concave on $\mathcal{X}$, and that the feasible set Ξ is convex, it is possible to design distributed algorithms for solving (1.2); see, for example, the results in [82, 107].

Notice that in the NUM problems introduced above we assumed the concavity of the utility function U, as well as, the convexity of the feasible set Ξ. These two features simplify their analysis and the design of either centralized or distributed algorithms converging to globally optimal solutions (under some assumptions). However, not all the resource allocation problems satisfy those strong assumptions. In particular, a great deal of applications in signal processing to communications violate them. For example, and of great interest to this work, is the so-called power allocation problem in communication systems adopting a dynamic spectrum management framework (see, Subsection 1.1.3), where the utility function corresponds to the sum of the transmission rates of all users (i.e. the so-called sum-rate) which is a nonconcave function. Furthermore, in the context of physical layer security (refer to Subsection 1.1.3), the utility function is nonconcave and also nondifferentiable. As a result, moving one step further, in this dissertation we study resource allocation problems that can be cast into the form (1.2), but the sum-utility function is nonconcave (refer to Chapter 3), or the objective function has the form of the sum of continuous max functions with a particular structure (refer to Chapter 4), giving rise to nondifferentiable and nonconcave utility functions.

1.1.2 Noncooperative Games

Since the introduction of game theory in [78] it has been adopted to analyze different problems in a wide variety of fields ranging from economics, psychology to engineering. Roughly speaking, game theory is a set of mathematical tools used to study and solve conflicts in scenarios in which different agents (the so-called players) make interactive decisions. We refer the reader to [80, 78] for a comprehensive treatment of this topic. Of particular interest to this dissertation is the area of game theory dealing with *noncooperative games*, that is, the case where the players make their individual decisions *selfishly* with the solely objective of optimizing their individual outcome. This outcome is generally measured via an objective function in the player's individual optimization problem. Thus, practically speaking, as opposed to the single optimization problem used in NUM to model a resource allocation problem, in the game theoretical framework, the game consists of multiple coupled optimization problems. It is worth mentioning that over the last decade or so, noncooperative game theory has been successfully employed to model different problems in signal processing and communications. This relevance comes from the fact that, in modern communication networks, distributed approaches are desirable since centralized ones may be too expensive or even infeasible to solve conflicts between their highly interactive entities.

In the rest of this section, we define and briefly introduce the following two types of games:

i) the Nash Equilibrium Problems (NEP); and

ii) the Generalized Nash Equilibrium Problems (GNEP).

1.1.2.1 Nash Equilibrium Problems

Let us formalize the concept of noncooperative games or also known as Nash equilibrium problems. Assume that there are I players denoted by the set $\mathcal{I} \triangleq \{1, \ldots, I\}$. Each player $i \in \mathcal{I}$ makes his decision on a n_i-dimensional real strategy vector $\mathbf{x}_i \in \mathcal{X}_i \subseteq \mathbb{R}^{n_i}$. Let $\mathcal{X} \triangleq \prod_{i=1}^{I} \mathcal{X}_i$ denote the admissible strategy set, and let $\mathcal{X}_{-i} \triangleq \prod_{j \neq i} \mathcal{X}_j$ be the strategy set of the rest of players distinct from player i. The outcome of each player i is measured through an utility function $U_i : \mathcal{X} \rightarrow \mathbb{R}$ that depends on the strategies of all the players in the game denoted by $\mathbf{x} \triangleq (\mathbf{x}_i, \mathbf{x}_{-i})$ where $\mathbf{x}_i \in \mathcal{X}_i$ and $\mathbf{x}_{-i} \in \mathcal{X}_{-i}$. Let $\mathbf{U}(\mathbf{x}) \triangleq (U_i(\mathbf{x}))_{i=1}^{I}$ denote the vector utility function. Then, the noncooperative game is the tuple $\mathcal{G} \triangleq (\mathcal{I}, \mathcal{X}, \mathbf{U})$, where every player $i \in \mathcal{I}$ aims to find the strategy vector $\mathbf{x}_i \in \mathcal{X}_i$, given the strategies of the rest of players $\mathbf{x}_{-i}$, that maximizes his utility function, i.e. anticipating $\mathbf{x}_{-i}$ each player i solves the following optimization problem

$$\begin{array}{ll} \underset{\mathbf{x}_i}{\text{maximize}} & U_i(\mathbf{x}_i, \mathbf{x}_{-i}) \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i. \end{array} \tag{1.3}$$

Note that the game $\mathcal{G}$ consists then in I coupled optimization problems of the form (1.3). It is worth stressing that in (1.3), the i-the player determines his strategy vector $\mathbf{x}_i$ by solving his corresponding optimization problem while considering the strategy vectors of the rest of players $\mathbf{x}_{-i}$ exogenous, that is fixed but arbitrary.

John Nash suggested in [76] that a solution to the game $\mathcal{G}$ is the feasible tuple $\mathbf{x}^\star \triangleq (\mathbf{x}_i^\star)_{i=1}^I$, the well-known Nash equilibrium (NE), with the characteristic that no player can improve his utility by unilaterally changing his decision, provided that the rest of players act according to it. The following definition introduces this concept formally.

Definition 1.1 (NE)**.** A strategy profile vector $\mathbf{x}^\star \triangleq (\mathbf{x}_i^\star)_{i=1}^I$ is a Nash equilibrium of the NEP $\mathcal{G} \triangleq (\mathcal{I}, \mathcal{X}, \mathbf{U})$ if the following holds for all $i \in \mathcal{I}$:

a) $\mathbf{x}_i^\star \in \mathcal{X}_i$; and,

b) $U_i(\mathbf{x}_i^\star, \mathbf{x}_{-i}^\star) \geq U_i(\mathbf{x}_i, \mathbf{x}_{-i}^\star)$ for all $\mathbf{x}_i \in \mathcal{X}_i$.

The existence of a NE for the game $\mathcal{G}$ is, in general, not guaranteed. The same holds with respect to its uniqueness. Both the NE existence and uniqueness for games in the general form $\mathcal{G}$ have been studied extensively in the related literature. The following two key observations have important implications on the NE existence and uniqueness.

- Given any $\mathbf{x}_{-i} \in \mathcal{X}_{-i}$, let $\mathcal{R}_i(\mathbf{x}_{-i})$ denote the best response of player i, that is, the set of optimal solutions to the i-th player optimization problem (1.3). Let $\mathcal{R}(\mathbf{x}) \triangleq \prod_{i=1}^I \mathcal{R}_i(\mathbf{x}_{-i})$ denote the best response map of the game $\mathcal{G}$. Then, it is clear that we can interpret the concept of NE introduced in Definition 1.1 as a fixed-point of the best response map $\mathcal{R}(\mathbf{x})$. More precisely, a strategy profile $\mathbf{x}^\star \in \mathcal{X}$ is a NE of the NEP $\mathcal{G}$ if and only if $\mathbf{x}^\star \in \mathcal{R}(\mathbf{x}^\star)$. This particular interpretation of the NE leads directly to some existence and uniqueness results for the NE of the game $\mathcal{G}$; these results are discussed below.

- A second approach that is useful to study the existence and uniqueness of the NE, and also helpful to derive provable convergent algorithms to compute it, is based on the reformulation of the game $\mathcal{G}$ as a variational inequality (VI). If we assume that for every $i \in \mathcal{I}$ it holds that: first, each set $\mathcal{X}_i$ is closed and convex; and, second, each utility function $U_i(\mathbf{x}_i, \mathbf{x}_{-i})$ is continuously differentiable in $\mathbf{x}$ and concave in $\mathbf{x}_i$ for every fixed $\mathbf{x}_{-i}$, then the game $\mathcal{G} \triangleq (\mathcal{I}, \mathcal{X}, \mathbf{U})$ is equivalent to the VI($\mathcal{X}, \mathbf{F}$), where $\mathbf{F} \triangleq (-\nabla_{\mathbf{x}} U_i(\mathbf{x}))_{i=1}^{I}$.

Among the different NE existence results, an important one capitalizes on the first observation above, i.e. on the interpretation of the NE concept as a fixed-point of the best response map, and then it applies the widely explored fixed-point theory to derive it. A second possible approach relies on the VI reformulation of the NEP explained above; we refer the reader to [29] for a more detailed insight. For our purpose, it is enough to recall the following well-known result guaranteeing the existence of a NE for the game $\mathcal{G}$; this is stated formally in the theorem below, whose proof follows from Kakutani's fixed-point theorem [28, Thm. 2.1.19].

Theorem 1.1 (NE existence)**.** Given the NEP $\mathcal{G} \triangleq (\mathcal{I}, \mathcal{X}, \mathbf{U})$, suppose that, for every $i \in \mathcal{I}$, the following two conditions hold

a) each player's strategy set $\mathcal{X}_i$ is convex and compact; and,

b) each player's utility function $U_i(\mathbf{x}_i, \mathbf{x}_{-i})$ is continuous in $\mathbf{x}$ and concave in $\mathbf{x}_i$ for every fixed $\mathbf{x}_{-i}$.

Then, the NEP has a nonempty solution set.

With respect to the uniqueness of the NE for the game $\mathcal{G}$, notice that, under the assumptions ensuring the equivalence between the game $\mathcal{G}$ and the VI($\mathcal{X}, \mathbf{F}$), it follows readily (from VI theory) that the NEP $\mathcal{G}$ admits a single solution if the map $\mathbf{F}$ is strongly monotone over $\mathcal{X}$; refer to [29] for sufficient conditions guaranteeing the required strong monotonicity of $\mathbf{F}$. Interestingly, the aforementioned condition is also sufficient to guarantee the convergence of distributed algorithms of the Gauss-Seidel or Jacobi type [12] to the unique NE of the game $\mathcal{G}$. We recall that, in Gauss-Seidel schemes, each player updates his strategy profile sequentially by solving the problem

(1.3) given the strategies of the rest of players; while in the Jacobi schemes, such updates are made in parallel. For the sake of clarity, we illustrate a Jacobi-type algorithm, where given $\mathbf{x}_{-i}^{\nu} \triangleq \left(\mathbf{x}_j^{\nu}\right)_{j \neq i} \in \mathcal{X}_{-i}$ each player $i \in \mathcal{I}$ computes

$$\mathbf{x}_i^{\nu+1} \in \underset{\mathbf{x}_i \in \mathcal{X}_i}{\operatorname{argmax}} \; U_i(\mathbf{x}_i, \mathbf{x}_{-i}^{\nu}).$$

It is worth mentioning that, a similar uniqueness result can be obtained if the best response map $\mathcal{R}(\mathbf{x})$ is a contraction in some norm; see, e.g., [28].

1.1.2.2 Generalized Nash Equilibrium Problems

So far we have restricted our discussion to games in which the coupling of the players' optimization problems occurs only at the level of their objective functions. However, there are situations in which the strategy set of every player depends on the strategies of the rivals i.e. the strategy set of user i depends on $\mathbf{x}_{-i}$, hence we denote this dependency explicitly by writing $\mathcal{X}_i(\mathbf{x}_{-i})$. This is indeed the case in the physical layer based security applications discussed in Chapters 3 and 4 of this dissertation. This extended notion of NEP is known as generalized Nash equilibrium problem (GNEP); we refer the interested reader to [27] for a detailed treatment of this topic.

Under the same setting of the NEP and using the same notation therein, each player i seeks the strategy profile $\mathbf{x}_i$ that maximizes his utility function, given the strategy of the others $\mathbf{x}_{-i}$, that is, every player $i \in \mathcal{I}$ aims to solve the following optimization problem

$$\begin{aligned} \underset{\mathbf{x}_i}{\text{maximize}} \quad & U_i(\mathbf{x}_i, \mathbf{x}_{-i}) \\ \text{subject to} \quad & \mathbf{x}_i \in \mathcal{X}_i(\mathbf{x}_{-i}). \end{aligned} \tag{1.4}$$

The next definition introduces a solution concept for the GNEP described above; as expected, this is a natural extension of that one in Definition 1.1.

Definition 1.2 (GNE)**.** A strategy profile vector $\mathbf{x}^{\star} \triangleq (\mathbf{x}_i^{\star})_{i \in \mathcal{I}}$ is a GNE of the GNEP $\mathcal{G} \triangleq (\mathcal{I}, \mathcal{X}(\mathbf{x}), \mathbf{U})$ if the following holds for all $i \in \mathcal{I}$:

a) $\mathbf{x}_i^{\star} \in \mathcal{X}_i(\mathbf{x}_{-i}^{\star})$; and

b) $U_i(\mathbf{x}_i^{\star}, \mathbf{x}_{-i}^{\star}) \geq U_i(\mathbf{x}_i, \mathbf{x}_{-i}^{\star})$ for all $\mathbf{x}_i \in \mathcal{X}_i(\mathbf{x}_{-i}^{\star})$.

The coupling in the feasible sets of the players introduced in the GNEP makes its solution significantly harder than the NEP. Similarly, the existence results for a GNE require, in general, more restrictive assumptions than the ones for the NE (see, e.g., Theorem 1.1). Since, the GNEP in its general form is almost intractable [98], different attempts have been proposed in the literature to derive results for instances of the GNEP with some particular structure; examples of those works are [31, 26].

In the applications section of Chapters 3 and 4, we study GNEPs where the objective function of the players' optimization problems is nonconcave and nondifferentiable. These features coupled with its generalized type make this sort of games challenging. However, by exploiting the structure of the problem of interest we are still able to analyze it and thus, derive iterative algorithms converging to relaxed equilibrium points of $\mathcal{G}$.

In the next subsection, we briefly introduce the context for the applications in signal processing to communication that are studied in this dissertation. A more detailed discussion of each of those applications is given in their respective chapters and in the cited references.

1.1.3 Spectrum Management

The demand experimented by either wired or wireless communication systems has increased dramatically in recent years. This proliferation of multiuser communication systems has motivated the study of different techniques to optimize their performance. Among them, the optimal allocation of the network resources is critical. For example, in wireless (e.g., ad-hoc networks) and wired (e.g., Digital Subscriber Lines−DSL)) systems, multiple users aim to communicate over a common medium causing *interference* among them. Hence, the overall system performance can be greatly affected due to the multiuser interference. These undesirable effects can be mitigated through proper spectrum management techniques.

As an example of the interference problem, let us consider the case of DSL, a broadband access technology, where multiple users obtain access to network services by communicating with a central node over the local telephone lines, thus sharing a common spectrum. The electromagnetic interference between the cables, known as crosstalk, may cause signal distortion. As a result, if the crosstalk is not mitigated, it can significantly limit the DSL system's performance. Indeed, in [120] the authors state that the crosstalk is possibly the major source of signal distortion in DSL. It is worth mentioning that a similar multiuser interference problem is faced by wireless networks. Consequently, it is evident that an adequate management of the spectrum is mandatory in order to avoid the degradation in the performance of these communication systems.

Basically, there are two different approaches for managing the spectrum, namely, the static and the dynamic techniques. Next, we provide an overview of both approaches.

Static Spectrum Management (SSM). SSM is the most basic and conventional spectrum management technique. For example, two well-known SSM approaches are: Frequency Division Multiple Access (FDMA) in which the network users share a spectrum divided into non-overlapping subchannels; and, similarly Time Division Multiple Access (TDMA), where the users transmit over non-overlapping time slots (see, e.g.,[110]). Notice that, since the users transmit over pre-assigned subchannels (or time slots), these cannot be used by another network entity when they are idle. Moreover, such fixed allocation policies may not be optimal. In the case of DSL systems, a common approach consists in allocating the spectrum based upon a worst-case scenario analysis; which may however reduce the achievable transmission rates. Hence, it follows that, the SSM techniques may lead to a conservative and restrictive network performance. The main reason of the inefficiency of the SSM is that it does not take into account the high variability or dynamism that every communication network faces.

Dynamic Spectrum Management (DSM). As a response to the suboptimal spectrum allocations that SSM techniques may produce, dynamic spectrum management approaches have been proposed as a viable solution to this problem; we refer the interested reader to [127] for a general survey on this topic, and to [69] for an analysis from an optimization point of view. The main idea behind DSM is to allocate the spectrum in such a way that it adapts to the characteristics of the network while achieving some network's performance metric. For example, the interference problem can be mitigated through *power control* by DSM techniques if the users in the communication system are allowed to allocate their available power budget dynamically across the entire shared spectrum, with the objective of maximizing the systems' throughput. Clearly, optimization techniques are required in order to obtain such optimal spectrum allocations that the DSM aims to find. The DSM problems can be formulated either as:

i) a single network utility maximization (NUM) problem (see, Subsection 1.1.1) [69]; or,

ii) as a noncooperative game (see, Subsection 1.1.2).

Several algorithms have been proposed in the related literature following the two aforementioned approaches; for example, we refer the reader to [45] for a complete discussion on DSM algorithms in DSL systems. To this end, it is worth emphasizing that, in this dissertation, we focus our attention on communication systems adopting a DSM framework. Moreover, in Chapter 2 we contrast the NUM and the game theoretical DSM approaches by analyzing the sum-utility obtained from a noncooperative game, when the power budget in the system is increased toward infinity.

Basically, the DSM algorithms can be categorized in the following two different classes:

- **Centralized DSM Algorithms.** Centralized schemes require coordination among the entities in the network achieved via a central node. This centralized authority is responsible of obtaining from the network nodes the information required to compute the optimal allocation of the spectrum, and then, communicate such an optimal policy to the corresponding nodes. In general, centralized schemes are computationally expensive, however the performance achieved by these schemes may outperform the obtained from distributed approaches, since the formers aim to compute globally optimal solutions. However, generally speaking, this approach may be intractable for large-scale networks. See, [45] for some examples.

- **Distributed DSM Algorithms.** In distributed schemes, there is no need of a central authority to optimally allocate the spectrum. Thus, the computations are decentralized across the network entities, and the level of coordination required among them is negligible. There are cases in which some limited signaling (in the form of message passing) between the nodes is needed by the algorithms. Distributed schemes are expected to be computationally less expensive than centralized algorithms, but at the cost of a reduction in the achieved system's performance. Due to their reduced complexity, distributed DSM algorithms are, in general, scalable and thus, suitable for communication systems with a large number of users.

The DSM based on the maximization of the network's utility function gives rise to a *centralized* planned solution or social optimum. Unfortunately, if we consider, for example, the single optimization problem formulation where the objective is to maximize the sum-rate subject to individual power constraints, such a program is in general nonconcave and it has been shown to be NP-hard [37]. Consequently, if devising DSM centralized algorithms is in jeopardy, the design of distributed schemes reaching the social optimum is even more challenging. However, different attempts have been introduced in the literature to find suboptimal solutions of the aforementioned problem; for example [65, 18, 19, 121] use a dual approach, while [44, 93] propose a sequential decomposition scheme and, more recently [94] introduces a parallel approach capitalizing on Successive Convex Approximation (SCA) techniques (see, e.g., [72, 90, 94, 4]) to reach a stationary solution of the NUM problem.

Under the same context, the DSM based on game theoretical models was first proposed by Yu et al. in [120] and has been extensively studied in the literature; we refer the reader to [98, 102, 20, 118, 68, 97, 96, 17, 54, 58], and the references therein for some relevant works. In this case, each network user is modeled as a selfish player that aims to maximize its information rate while considering the power allocation of the others as measurable noise. The resulting individual players' optimization problem are concave, giving rise to a Nash equilibrium problem. Contrary to the NUM approach described above, the NEP leads directly to a *distributed* implementation obtained via the well-known Iterative Waterfilling Algorithm (IWFA), see, e.g., [95]. It is important to mention that the convergence of the IWFA to a NE of the so-called power allocation game is not always guaranteed, see, e.g., [68, 96, 17] for some sufficient conditions.

Game theoretical models have been employed successfully to allocate the spectrum dynamically in different communication models. In this dissertation, we analyze some of those existing system designs and also introduce novel models, such as the DSM in networks implementing physical layer based security (refer to Chapters 3 and 4). In order to set the context of the research proposed in this dissertation, we briefly revise some of the most relevant system designs introduced in the literature.

In [86] the authors consider a noncooperative game in which each network user is modeled as a player that aims to find the power allocation that satisfies a desired information rate (the so-called Quality of Service − QoS) with the minimum possible power. Notice that, in this case, the objective is to minimize the power rather than maximizing the transmission rate as in the power allocation game described above. The idea of introducing QoS constraints arises as a need of overcoming the possibly unfair power allocations obtained from the classical approach that tend to assign higher data rates to those users with "better" channels. A totally different situation that can happen in a communication network is the presence of malicious nodes that aim to somehow disturb the normal operation of the system. The authors of [33] studied such situation by introducing a power allocation game where an antagonistic player (or also known as jammer) aims to disrupt the network performance by minimizing the utility of the entire system.

The increasing demand of wireless services during recent years, along with fixed spectrum allocation policies have brought as a negative consequence the scarcity of the radio spectrum. Cognitive Radio (CR) has been introduced as a possible solution to this problem; see, e.g., [99, 38] for two surveys on this topic. Roughly speaking, the CR paradigm introduces a hierarchical structure in the communication system where the users can be classified into: (i) primary users or the legacy spectrum holders, and (ii) secondary users who are unlicensed spectrum holders with the ability to access the licensed spectrum with the restriction of not degrading the primary users' quality of service. As expected, game theoretical models have been considered in order to devise distributed resource allocation policies in this context. Among the different works in this area, consider for example [87], where the optimal power allocation of the secondary users is obtained via a noncooperative game, in which a pricing mechanism is used to handle the coupling among the variables of all the players that the primary users' interference constraints impose. It is worth mentioning that, most of the analysis of this CR system design relies on VI theory. A more complex CR system design was considered in [85, 100]. In particular, the cited references take into account the fact that the access to the licensed spectrum from the secondary users depends on their capabilities to detect the so-called "spectrum holes" (i.e. unused spectrum slots). This calls for a joint optimization of the power allocation along with these detection parameters, giving rise to a noncooperative and nonconvex game with side constraints. Interestingly, this game motivated the work in [84] as a response to the lack of mathematical tools available in the literature to analyze the aforementioned problem. In the cited work, a new concept of equilibrium was introduced, namely, the *quasi-Nash Equilibrium* (QNE). Essentially, a QNE is a solution of the VI obtained by aggregating the first order optimality conditions of the players' optimization problems while the convex constraints are retained in the defining set of the VI. Whenever a NE of the game exits, it must be a QNE under some constraint qualifications. This new concept opened the path to analyze the nonconvex CR game, proving that such a game always admits a QNE. In this dissertation, we use and extend the concept of QNE to encompass games where the objective functions of the player's optimization problems are besides nonconvex also nondifferentiable.

With most of our daily transactions being done through the Internet, security is a major concern in today's communication networks. Different from current cryptographic techniques, used to guarantee security among the entities in the system, the objective of physical layer based security is, as expected from its name, to exploit the physical characteristics of the communication channel in pursuance of secure transmissions; see, e.g., [63, 61, 47] for recent surveys on this topic. Since the seminal work of Aaron Wyner in [116], the idea of physical layer security has been considered as a promising technique to provide secure communications. Of interest to our work, is the so-called Cooperative Jamming (CJ) paradigm (see, e.g., [24, 39, 59]), in which some particular nodes in the network (known as friendly jammers) are introduced into the communication system with the objective of generating judicious interference in order to improve the *secrecy rate*, that is, roughly speaking, the non-zero rate at which the legitimate users can communicate in a secure way.

Again, different game theoretical models have been employed to dynamically allocate the spectrum in this context, among them we can mention those in [36, 35, 125, 115, 104]. There are two major weaknesses that all of these works exhibit, first the network models analyzed are very simplistic, for example, they are composed of a single transmitter-receiver pair and multiple friendly jammers or multiple transmitter-receiver pairs but only one friendly jammer; and second, those models ignore a nondifferentiability issue that is inherent to every secrecy rate problem. In this dissertation, we overcome both issues by introducing more general network models, and by analyzing them carefully taking into account the nondifferentiability problem.

To conclude this section, it is important to emphasize that the resource allocation problems considered in this dissertation focus on communication systems adopting the DSM framework in two different contexts: (i) in the most general setting of mitigating the interference problem through power control; and, (ii) in the the emerging field of physical layer based security, where, in contrast to the former case, the interference may be beneficial to secure the transmissions between legitimate parties provided that it is between acceptable thresholds and, of course, properly managed. Although, we center the discussion in these applications, it is worth remarking that our techniques and results can be readily applied to other resource allocation problems in multiuser systems.

1.2 Research Synopsis

This dissertation is divided into three main chapters. These chapters are independent, thus, the reader can read them in any order after going over this introductory section. Let us briefly discuss each of them.

In Chapter 2, we present an analysis based on the Linear Complementarity Problem (LCP) theory [21] of the dependance on the power budget of the system sum-rate obtained from the outcome of a NEP used to dynamically allocate the spectrum. More precisely, in favor of obtaining some insight for the situation of relatively large but finite power budgets, we study the case when the budget is increased toward infinity. This study led us to draw some conclusions on the efficiency of the NE under the unbounded budget setting. Interestingly, we observed that the system sum-rate of NE could be finite even in the presence of an infinite power budget, contrary to the unbounded sum-rate obtained when a cooperative approach is used to allocate the spectrum. Hence, the unboundedness of the power resource may not be enough to overcome the users' selfishness. Furthermore, this research also led us to find, a particular case for which the presence of the well-known Braess-type paradox [14] is ruled out in the interference channels.

In Chapter 3 we propose a decomposition method for the minimization of a nonseparable sum-utility function of the Difference of Convex (DC) type, subject to a set of convex coupling constraints. One of the main contributions of this chapter is the design of a class of (inexact) distributed algorithms with provable convergence to stationary points of the DC problem. The proposed algorithms are based on SCA-techniques, that is, a suitable convexified version of the original problem is solved at each iteration. Among the different possible applications of our algorithms, we apply them to the solution of the following two problems: (i) the sum-rate maximization problem in CR MIMO (Multiple-Input-Multiple-Output) systems, providing for the first time a provable convergent algorithm to address such resource allocation problem; and (ii) to a power allocation game in the context of physical layer security. This game faces the following difficulties: first, the objective functions of the users' optimization problems are nonconcave and nondifferentiable; and second, there are also constraints coupling the strategies of the users. In order to provide a rigorous mathematical analysis of this game, we introduce a new relaxed equilibrium concept based on directional derivatives, termed B-Quasi Generalized Nash Equilibrium (B-QGNE), which is shown to be reachable applying the DC framework described above. Numerical results validating our theory and evaluating the performance of the proposed algorithms are also provided.

Chapter 4 introduces a multiuser optimization problem where the nonseparable utility function is the sum of continuous max functions with a particular structure. In order to develop provable convergent algorithms to address this problem, we reformulate it in such a way that the obtained problems overcome some of the main difficulties present in the original one, while still being associated with it. We also study carefully the connections between those "simpler" reformulations and the original problem. One of the mentioned reformulations arises from a (nontrivial) DC decomposition of the objective function of such a problem, while the second approach is based on a smooth reformulation of the problem. The former approach lead us to apply the well-known DCA (Difference of Convex Algorithm) [108, 109, 5, 122, 53, 103], and for the latter approach, SCA techniques are applicable. Our main interest in studying this particular multiuser optimization problem lies in deriving provable convergent algorithms for the allocation of resources in communication systems implementing physical layer based security, where the users communicate over *multiple* subchannels, this last feature is the key distinction from the system model considered in Chapter 3 and from those studied in the literature. The proposed algorithms are tested numerically under different settings for the aforementioned application.

Finally, Chapter 5 summarizes the overall discussion and draws the main conclusion. Future research topics are discussed at the end of each chapter.

The overall main and unique contributions of this dissertation can be summarized as follows:

1. A Linear Complementarity Problem based framework to analyze the asymptotic behavior of the utility of a communication system, when a noncooperative game is used to allocate the spectrum and when the users' power budgets are increased toward infinity.

2. A class of distributed algorithms with provable convergence to stationary solutions of multiuser optimization problems with smooth, nonconvex and nonseparable objective function subject to convex coupling constraints.

3. A rigorous treatment of a game (of the generalized type) arising in a novel resource allocation problem in the field of physical layer based security, where the players' optimization problems are characterized to have a nonconvex and nondifferentiable objective function.

4. A set of provable convergent algorithms for addressing a sum-utility maximization problem where the objective function is the sum of continuous max functions of a particular kind.

5. The development of (distributed) algorithms for the power allocation problem in a multi-user, multi-friendly jammers and multi (orthogonal)

subchannels communication system securing its transmission through physical layer, ranging from the secrecy rate maximization problem to the max-min fairness system design. Further extensions include the case of a MIMO communication system.

This dissertation extends the current results in the literature in two major areas: first, in the field of optimization techniques applied to address multiuser resource allocation problems in either a centralized or distributed fashion; and second, in the domain of signal processing, more precisely in the power allocation problem faced by communication systems under DSM. For the sake of clarity, Figure 1.1 illustrates the overall structure of this dissertation. This figure also highlights how each of the chapters outlined above extend the current results in the related literature.

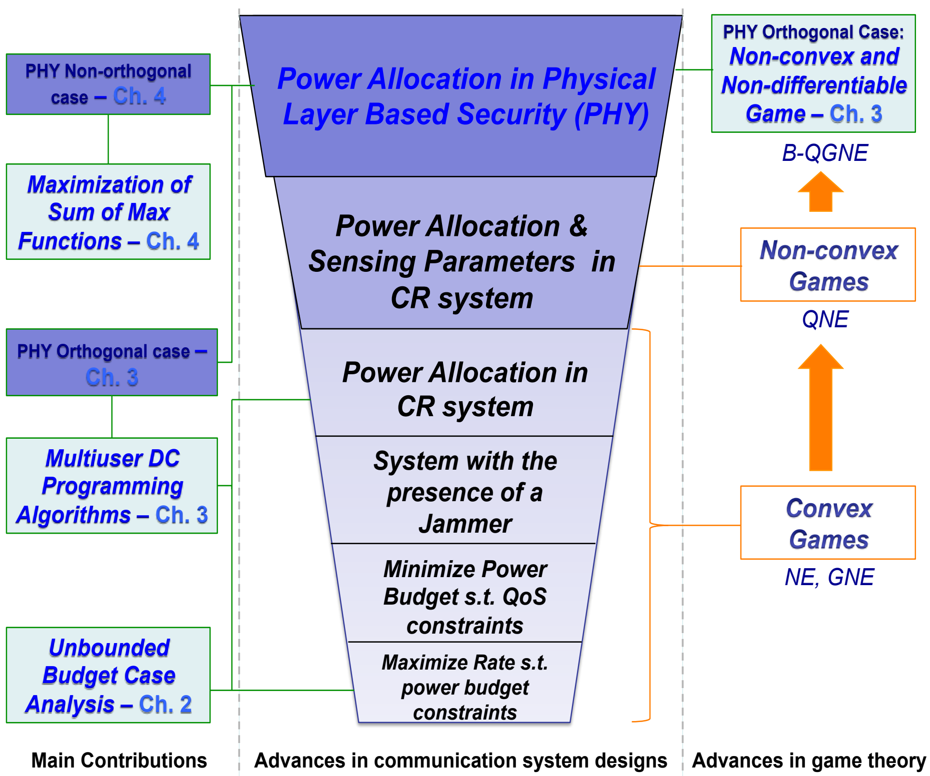

Figure 1.1: Main contributions of this dissertation highlighted by chapter.

Chapter 2

Dependence on the Power Budget of the System Sum-Rate of Nash Equilibria[1]

2.1 Introduction

With the increasing popularity of multiuser communication systems, their optimal performance is becoming a very important issue. In wired systems, such as Digital Subscriber Lines (DSL), and wireless systems, like ad-hoc networks, multiple users try to communicate over a common medium causing *interference* to each other. In this chapter, we model these interfering networks as Gaussian Interference Channels (IC). Clearly, an adequate management of the interference is required to avoid the degradation of the overall system's performance.

In a dynamic spectrum management (DSM) framework, the aforementioned problem has been addressed through *power control* using two approaches. First, an approach based on the maximization of the so-called "social-function" i.e. the sum of the transmission rates of all the users across all subchannels (hereafter and for shortness, we will refer to it as *system sum-rate*) subject to individual power constraints (see, e.g., [65, 18, 19, 121]). Unfortunately, this approach leads to a *centralized* planned solution, and the resulting optimization problem is in general nonconvex and has been shown to be NP-hard [37]. A different solution approach that overcomes these difficulties, first proposed by Yu et al. in [120] and further analyzed in [102, 20, 118, 68, 97, 96, 17, 54, 58, 98], is based on game theoretical models. In this case, each network user acts as a selfish player that aims to maximize its information rate while considering the power allocation of the others as measurable noise; giving rise to a noncooperative (Nash) game [76]. This approach leads directly to a *distributed* implementation obtained

[1] This chapter is adapted from a manuscript being prepared for submission (Co-author: Jong-Shi Pang).

via the well-known Iterative Waterfilling Algorithm (IWFA), see, e.g., [95]. However, the convergence of the IWFA to a Nash Equilibrium (NE) of the so-called power allocation game is guaranteed only under some conditions, see, e.g., [68, 96, 17].

Based on the centralized and distributed solution schemes for the dynamic power allocation problem, the main objective of this chapter is to study the dependence on the power budget of the system sum-rate obtained from the Nash equilibrium (NE). Ideally, if the power budget in the communication system is increased, we would expect the system sum-rate of NE to increase as well. However, according to [67, 3, 2] and further computational results, this is not necessarily the case. Therefore, a *Braess-type paradox* may occur in the Gaussian IC. Such a paradox, introduced in [14] in the context of a traffic model, asserts that there are situations in which, if the users in the system act selfishly, then an increase in the power budget do not necessarily lead to an increase in the overall system's performance. In the same vein, if the power budget is unbounded we would expect that the system sum-rate of NE to approach infinity as well. Again, a numerical example in [25] and our simulation results show that this expectation is not always realized.

Motivated by the aforementioned scenarios, in this chapter, we focus on the situation when the users are endowed with the *same* power budget that tends to infinity; the cases when the users have budgets that tend to infinity disproportionally or when some users have bounded budgets while others do not, can be similarly analyzed but is omitted. In particular, since the optimum centralized system sum-rate must be equal to infinity in this case, we therefore study the limit of the system sum-rate of NE as the players' budgets become unbounded. Although the power budget is a bounded resource, we consider the unbounded case in favor of studying situations with finite but arbitrarily large power budget, with the former limiting case being an approximation of the latter finite case. Furthermore, our interest in studying this topic comes from the following two interesting cases that could happen for such a limit: (a) it is bounded, and (b) it is infinite. The consequence of case (a) is that it does not matter how much power budget the users in the network have, the system sum-rate of NE will always be finite; thus the endowment of infinite power budgets for all users may not be enough to overcome their selfishness. In case (b), both the NE-based system sum-rate and the optimum centralized one will approach infinity as the power budget is increased toward infinity.

Different approaches have been proposed in the literature to study the *efficiency* of the NE in the context of interference channels. One approach is based on the concepts of price of anarchy (PoA) [52] and price of stability (PoS) [6], defined as ratios between the optimal centralized solution and the worst/best NE respectively; see, e.g., [73, 71] where the case of two users is studied. More related to the present work, the paper [73] studies the efficiency of the NE at high Signal to Noise Ratio (SNR) for the 2-users case with the aid of the so-called *high-SNR measures* i.e. via the concepts of high-SNR slope (see, e.g., [110, 46]) and high-SNR power offset [64]. A more recent work that also touches the efficiency of the NE is [10], where the approach is based on a (approximately) characterization of the "NE region" of the 2-user (Gaussian) linear-deterministic IC [15]. Deviating from these approaches and without restricting in the number of users and subchannels, we capitalize on the Linear Complementarity Problem (LCP) theory [21] to draw conclusions on the efficiency of the NE under the unbounded budget setting described above. As a side note, LCP concepts will be freely employed in the discussion, which can be found in the cited monograph.

In order to achieve the objectives stated above, we follow the approach outlined next. First, we formulate an optimization problem that seeks to maximize the system sum-rate over the set of NE, which, as shown in [68], corresponds to the set of solutions to a certain LCP. For shortness this optimization problem will be denoted by $\mathrm{MSSR}_{\mathrm{NE}}$ (Maximum System Sum-Rate over the set of Nash Equilibria). Second, to study the limiting behavior of this problem, as the power budget increases toward infinity, we introduce an auxiliary optimization problem that is obtained by homogenizing the LCP constraints of the $\mathrm{MSSR}_{\mathrm{NE}}$; which we denote by $\mathrm{HMSSR}_{\mathrm{NE}}$. Finally, we develop sufficient conditions for the optimum objective values of these two problems to coincide when the power budget goes to infinity.

The approach described above gives rise to our first contribution, a framework for analyzing the dependance on the (unbounded) power budget of the NE-based system sum-rate. Capitalizing on the $\mathrm{HMSSR}_{\mathrm{NE}}$ problem, a second contribution is the characterization of the maximum attainable system sum-rate of NE, and that of the NE tuple that yields an infinite system sum-rate of NE when the power budget is unbounded. Furthermore, we develop a (theoretical) constructive test for verifying such condition. This analysis gives rise to some special cases that yield sharper results. A third contribution lies in one of such cases, which we call the case of *equi-noisy channels*, that eliminates the presence of the Braess-type paradox in the Gaussian IC. We also provide sufficient conditions under which the set of NE is finite. Finally, our numerical results show that it is possible to observe an asymptotically *finite* system sum-rate of NE (for different network's parameters) as the system's power budget is increased toward infinity. Thus implying that the cooperative approach outperforms the noncooperative one even when the power is an unlimited resource. To the best of our knowledge, this chapter proposes a novel line of asymptotic analysis for multiuser communication systems over Gaussian IC built upon optimization theory. In addition, this work introduces a new way of assessing the performance of these systems when a game-theoretical approach is employed to dynamically allocate the spectrum.

The rest of this chapter is organized as follows. Section 2.2 introduces the system model and the MSSR_{NE} problem. Section 2.3 presents some numerical examples that motivated the present work. In Section 2.4 we introduce the proposed framework for analyzing the system sum-rate of NE as the power budget goes to infinity. In Section 2.5 we derive a necessary and sufficient condition on the NE tuple that yields an infinite system sum-rate for the unbounded power budget case. Section 2.6 introduces some special cases and Section 2.7 contrasts our findings with some numerical simulations. Finally, Section 2.8 draws some conclusions and presents future research topics.

2.2 System Model and Problem Formulation

In this section, we describe the system model including the underlying assumptions and notation to be used throughout this chapter. In Subsection 2.2.2, we formulate the problem to be analyzed in detail during the discussion.

2.2.1 System model

We consider a Q-user N-parallel Gaussian interference channel. In this system model, each user $q = 1, \ldots, Q$ is a single transmitter-receiver pair, where each transmitter aims to communicate with its corresponding receiver over the shared spectrum composed of the subchannels $k = 1, \ldots, N$. Under basic information theoretical assumptions and invoking the capacity expression for the single user Gaussian channel (achievable using random Gaussian codes from all the users), the maximum information rate on link q for a specific power allocation profile is given by [22]

$$R_q(\mathbf{p}_q, \mathbf{p}_{-q}) \triangleq \sum_{k=1}^{N} \log \left(1 + \frac{|h_{qq}(k)|^2 \, p_q(k)}{N_q^2(k) + \sum_{r \neq q} |h_{rq}(k)|^2 p_r(k)} \right), \tag{2.1}$$

where $|h_{rq}(k)|^2$ is the power gain of the channel between transmitter r and receiver q over subchannel k; $N_q^2(k)$ is the variance of the thermal noise over subchannel k at receiver q; the vector $\mathbf{p}_q \triangleq (p_q(k))_{k=1}^N$ represents the power allocation strategy of user q across all subchannels, and the vector $\mathbf{p}_{-q} \triangleq (\mathbf{p}_r)_{r \neq q}$ stands for the power allocation strategies of the rest of users in the network. By letting $H_{rq}(k) \triangleq |h_{rq}(k)|^2 / |h_{qq}(k)|^2$ and $\sigma_q^2(k) \triangleq N_q^2(k) / |h_{qq}(k)|^2$ for all $q, r = 1, \ldots, Q$ and all $k = 1, \ldots, N$, (2.1) can be rewritten as follows

$$R_q(\mathbf{p}_q, \mathbf{p}_{-q}) = \sum_{k=1}^{N} \log \left(1 + \frac{p_q(k)}{\sigma_q^2(k) + \sum_{r \neq q} H_{rq}(k) \, p_r(k)} \right). \tag{2.2}$$

For the sake of simplicity, throughout the rest of the discussion, we will use the normalized form of $R_q(\mathbf{p}_q, \mathbf{p}_{-q})$ given in (2.2).

2.2.2 Problem Formulation

Before introducing the main problem that will be studied in the subsequent analysis, we briefly revisit the two basic approaches that have been proposed in the literature to address the dynamic power allocation problem in a communication system where users share a common spectrum; namely, the centralized and decentralized solution schemes.

- *Centralized Approach*: it consists in optimizing the system's utility function i.e. the sum of the information rate of every user across all subchannels (system sum-rate) subject to individual power constraints (see, e.g., [65, 18, 19, 121, 37]). Assuming each user q has a fixed power budget $B > 0$ (and possibly some mask constraints), the sum-rate maximization problem is given by

$$\begin{aligned} R_{\boldsymbol{\sigma}}^{\text{sys}}(B) \triangleq \; & \underset{(\mathbf{p}_q)_{q=1}^{Q} \geq \mathbf{0}}{\text{maximum}} \quad \sum_{q=1}^{Q} R_q(\mathbf{p}_q, \mathbf{p}_{-q}) \\ & \text{subject to} \quad \sum_{k=1}^{N} p_q(k) \leq B \quad \forall\, q = 1, \ldots, Q. \end{aligned} \tag{2.3}$$

Remark 2.1 (On the System Problem (2.3))**.** It is worth stressing that $R_{\boldsymbol{\sigma}}^{\text{sys}}(B)$ denotes the optimal objective value of the single multiuser optimization problem (2.3) for a power budget B (where the dependance is made explicit) and for any tuple of noise variances $\boldsymbol{\sigma} \triangleq \left(\boldsymbol{\sigma}_q \triangleq \left(\sigma_q^2(k)\right)_{k=1}^{N}\right)_{q=1}^{Q}$. Hence, we refer to $R_{\boldsymbol{\sigma}}^{\text{sys}}(B)$ as the *centralized maximum system sum-rate*. Additionally, it is not difficult to show that $\lim_{B\to\infty} R_{\boldsymbol{\sigma}}^{\text{sys}}(B) = \infty$ for any tuple $\boldsymbol{\sigma}$. Hence, under the centralized regime, the maximum system sum-rate must be unbounded as the users' power budgets tend to infinity. In this "centralized regime", different approaches have been proposed in the literature to compute stationary solutions of the nonconvex problem (2.3) in a *distributed* way with some signaling among the users, see, e.g., [94] for a recent work.

- *Decentralized Approach*: it is based on noncooperative game theory (see, e.g., [120, 102, 20, 118, 68, 97, 96, 17, 54, 58, 98]). In this approach, each user maximizes its own information rate by considering the power allocation of the rival users as measurable Gaussian noise; thus, leading directly to a distributed implementation where only local channel state information is required by each user. More precisely, assuming each user q has a given power budget $B > 0$ (and possibly some mask constraints), the objective of user q is to maximize its information rate $R_q(\mathbf{p}_q, \mathbf{p}_{-q})$; i.e. anticipating $\mathbf{p}_{-q}$ each user $q = 1, \ldots, Q$ aims to solve the convex optimization problem

$$\begin{array}{ll} \underset{\mathbf{p}_q \geq \mathbf{0}}{\text{maximize}} & R_q(\mathbf{p}_q, \mathbf{p}_{-q}) \\ \text{subject to} & \displaystyle\sum_{k=1}^{N} p_q(k) \;\leq\; B. \end{array} \tag{2.4}$$

Remark 2.2 (Equivalent LCP Formulation)**.** In [68, Prop. 1], the authors established an equivalent LCP formulation for the noncooperative (Nash) game where each user q solves (2.4). We remark that in the cited proposition each player's problem includes mask constraints, however the aforementioned result follows readily by setting those masks equal to infinity. Let λ_q be the Lagrange multiplier corresponding to the linear constraint in (2.4). Then, based on the Karush-Kuhn-Tucker (KKT) optimality conditions for each user's optimization problem (2.4), the equivalent LCP formulation of the mentioned game is given by

$$\text{LCP}(\mathbf{b}, \mathbf{M})\text{:} \begin{cases} \forall\, q = 1, \ldots, Q \text{ and } \forall\, k = 1, \ldots, N & \\ 0 \leq \sigma_q^2(k) + \displaystyle\sum_{r=1}^{Q} H_{rq}(k) p_r(k) - \lambda_q & \perp \;\; p_q(k) \geq 0 \\ 0 \leq -B + \displaystyle\sum_{k=1}^{N} p_q(k) & \perp \;\; \lambda_q \geq 0. \end{cases} \tag{2.5}$$

In (2.5) the complementarity notation $0 \leq a \perp b \geq 0$ for scalars (or vectors) a and b means $a \cdot b = 0$, $a \geq 0$ and $b \geq 0$. For shortness, the LCP in (2.5) will be denoted by the pair $(\mathbf{b}, \mathbf{M})$:

$$\mathbf{0} \;\leq\; \mathbf{z} \quad \perp \quad \mathbf{M}\,\mathbf{z} + \mathbf{b} \;\geq\; \mathbf{0} \tag{2.6}$$

where

$$\mathbf{z} \triangleq \begin{pmatrix} \mathbf{p} \triangleq (\,\mathbf{p}_q)_{q=1}^{Q} \\ \boldsymbol{\lambda} \triangleq (\,\lambda_q\,)_{q=1}^{Q} \end{pmatrix} \in \mathbb{R}^{\mathbf{N}}$$

$$\mathbf{b} \triangleq \begin{pmatrix} \boldsymbol{\sigma} \\ -B\,\mathbf{1}_Q \end{pmatrix} \in \mathbb{R}^{\mathbf{N}}, \tag{2.7}$$

and the matrix $\mathbf{M} \in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$ is given by

$$\mathbf{M} \triangleq \left[\begin{array}{cccc|cccc} \mathbf{I}_N & \text{Diag}(\mathbf{H}_{12}) & \cdots & \text{Diag}(\mathbf{H}_{1Q}) & -\mathbf{1}_N & & & \\ \text{Diag}(\mathbf{H}_{21}) & \mathbf{I}_N & \cdots & \text{Diag}(\mathbf{H}_{2Q}) & & -\mathbf{1}_N & & \\ \vdots & \vdots & \ddots & \vdots & & & \ddots & \\ \text{Diag}(\mathbf{H}_{Q1}) & \text{Diag}(\mathbf{H}_{Q2}) & \cdots & \mathbf{I}_N & & & & -\mathbf{1}_N \\ \hline \mathbf{1}_N^T & & & & & & & \\ & \mathbf{1}_N^T & & & & & & \\ & & \ddots & & & & & \\ & & & \mathbf{1}_N^T & & & & \end{array}\right], \quad (2.8)$$

where $\mathbf{I}_N$ is the identity matrix of order N, $\mathbf{1}_N$ is the N-dimensional vector of all ones, $\mathbf{H}_{qr} \triangleq (H_{rq}(k))_{k=1}^N$ for $q \neq r$, $\text{Diag}(\mathbf{H}_{qr})$ is the diagonal matrix whose entries are the components of $\mathbf{H}_{qr}$, and the dimension $\mathbf{N} \triangleq NQ + Q$.

To conclude this section, we introduce an optimization problem that is the cornerstone of the analysis to be developed in the following sections. Namely, for a given power budget $B > 0$ and a tuple $\boldsymbol{\sigma}$ of noise variances, we define the following optimization problem

$$\text{MSSR}_{\text{NE}}: \begin{cases} R_{\boldsymbol{\sigma}}^{\text{NE}}(B) \triangleq \underset{p_q(k),\lambda_q}{\text{maximum}} \displaystyle\sum_{q=1}^{Q} R_q(\mathbf{p}_q, \mathbf{p}_{-q}) \\ \text{subject to } \forall\, q = 1, \ldots, Q \text{ and } \forall\, k = 1, \ldots, N: \\ 0 \leq \sigma_q^2(k) + \displaystyle\sum_{r=1}^{Q} H_{rq}(k) p_r(k) - \lambda_q \quad \perp \quad p_q(k) \geq 0 \\ 0 \leq -B + \displaystyle\sum_{k=1}^{N} p_q(k) \quad \perp \quad \lambda_q \geq 0. \end{cases} \quad (2.9)$$

Notice that the constraints of the maximization problem above correspond to the LCP$(\mathbf{b}, \mathbf{M})$. Since the solutions of the LCP$(\mathbf{b}, \mathbf{M})$ correspond to the Nash equilibria set of the game whose q-th player optimization problem is (2.4), the problem (2.9) is thus to seek a NE that maximizes the system sum-rate. Thus, we refer to $R_{\boldsymbol{\sigma}}^{\text{NE}}(B)$ as the *maximum system sum-rate of*

NE. Notice that, the dependence of the system sum-rate on the power budget B is made explicit by defining the function $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$. The superscript NE is attached to $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ to distinguish it from the centralized maximum system sum-rate $R_{\boldsymbol{\sigma}}^{\mathrm{sys}}(B)$ [cf. (2.3)]. Subsequently, we will be interested in a homogenized version of (2.9) where the noise variances $\sigma_q^2(k)$ are set equal to zero for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$; thus we also attach the subscript $\boldsymbol{\sigma}$ in $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ to denote the role of $\boldsymbol{\sigma}$ in the problem (2.9).

Remark 2.3 (About the $\mathrm{MSSR_{NE}}$)**.** The $\mathrm{MSSR_{NE}}$ problem belongs to the class of MPCC - *Mathematical Programs with Linear Complementarity Constraints* [66], that has been studied extensively in the mathematical programming literature. In this work, we are concerned mainly with a qualitative property of the $\mathrm{MSSR_{NE}}$, namely, whether its optimum objective value will be unbounded as the users' power budgets tend to infinity. It is important to mention that, in the case of a monotone game, i.e., when the resulting $\mathrm{LCP}(\mathbf{b}, \mathbf{M})$ is of the positive semidefinite type, the distributed algorithm developed in [30] can be applied to compute a stationary solution, albeit not necessarily a globally optimal solution of the $\mathrm{MSSR_{NE}}$. Details of these algorithmic issues for solving (2.9) are outside the scope of this study.

2.3 Motivating Examples

In this section, we introduce two numerical examples that motivated most of the analysis that is presented in the forthcoming sections. For the sake of clarity, we use these examples to introduce and illustrate the main questions that this chapter aims to answer.

Motivating Example 2.1. This numerical example, due to Z.Q. Luo in [67], considers the power allocation game where each user solves (2.4) for a system composed of $Q = 2$ users and $N = 2$ subchannels, with channel gain coefficients $H_{12}(k) = H_{21}(k) = 1$, and noise variances $\sigma_1^2(k) = \sigma_2^2(k) = 1$ for $k = 1, 2$. The relevance of this example is summarized in Table 2.1, from which it can be observed that when the power budget is increased, then the system sum-rate of NE decreases; thus showing the presence of the Braess-type paradox in the IC. The key point lies in the choice of the (non-unique)

NE tuple. In the first situation, the author chooses an orthogonal power allocation profile; while, in the second case, the sum-rate decreases when a uniform power allocation is used.

Table 2.1: The Braess-type paradox is present in the system described in Motivating Example 2.1.

Power Budget	NE tuple $(\mathbf{p}_1, \mathbf{p}_2)$	System Sum-Rate of NE (b/cu)
$B = 1.5$	$\mathbf{p}_1 = (1.5, 0)$, $\mathbf{p}_2 = (0, 1.5)$	2.64
$B = 2$	$\mathbf{p}_1 = (1, 1)$, $\mathbf{p}_2 = (1, 1)$	2.34

Based on the previous observations, a natural question to ask is: if among the NE tuples in the power allocation game, we consider *only* those that *maximize* the system sum-rate, is there any particular case that rules out the presence of the Braess-type paradox in the IC? The answer to this question is positive and it is analyzed in detail in Subsection 2.6.2, where we introduce the case of *equi-noisy* channels that eliminates the presence of such a paradox in this class of games.

Motivating Example 2.2. Lets examine a system composed of $Q = 2$ users and $N = 10$ subchannels under two different sets of channel realizations:

(a) Consider the flat-fading example introduced in [25], where the channel gain coefficients are $H_{12}(k) = H_{21}(k) = 1/4$, and the noise variances are $\sigma_1^2(k) = \sigma_2^2(k) = 1$ for $k = 1, \ldots, N$. It is not difficult to show that the unique NE for the power allocation game (2.4) corresponds to the uniform power profile $p_q(k) = \frac{B}{N}$ for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$. The authors of [25] observed that as $B \to \infty$, the transmission rates for each user $q = 1, 2$ approach $N \log_2(5)$, thus, the system sum-rate of NE approaches $2N \log_2(5)$.

(b) In contrast with the previous case, consider a more general setting with noise variances $\sigma_1^2(k) = \sigma_2^2(k) = 0.5$ for $k = 1, \ldots, N$ and channels simulated as in Section 2.7 - Example 2.3 with path loss exponent $\gamma = 2$ and normalized interlink distance $d_{rq}/d_{qq} = 3$.

The numerical results for both settings are summarized in Figure 2.1; where, by increasing the power budget B (for all users) from -10dBm to 60dBm, we can distinguish two cases. As noted in [25], Figure 2.1(a) shows that $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ exhibits a finite asymptotic behavior i.e. there is a point from which increasing the power budget does not translate in any performance improvement. On the contrary, Figure 2.1(b) illustrates that as the power budget is increased so does $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$. It is important to emphasize that, incidentally, the behaviors illustrated in this example were observed in many numerical simulations neither restricted to the flat-fading case, nor to number of users and subchannels (see, Section 2.7).

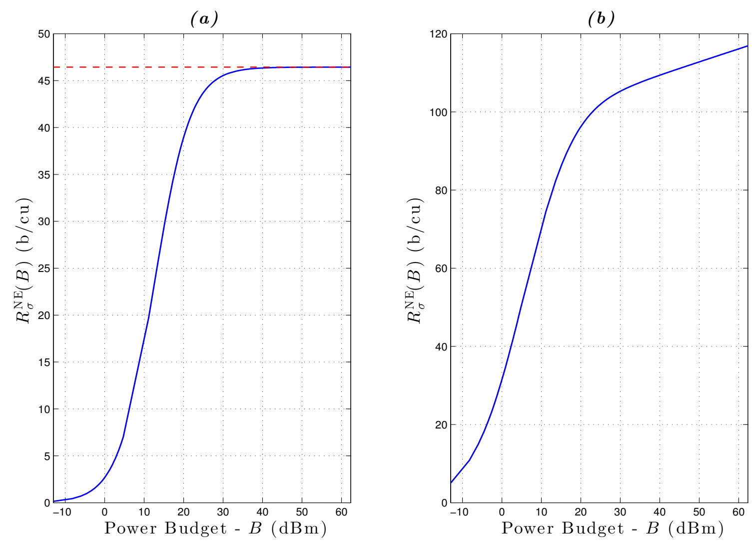

Figure 2.1: Maximum system sum-rate of NE ($R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$) versus power budget ($B$) in a system composed of $Q = 2$ users and $N = 10$ subchannels for two different sets of channel realizations (a) and (b) described in Motivating Example 2.2.

The relevance of the two cases depicted in the previous example lies in the following implications on the efficiency of the NE. The first case (in Figure 2.1(a)) suggests that $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) < \infty$ as $B \to \infty$, in other words, the unboundedness of the power budget is not enough to compensate for the selfishness of the players in the game. The second case (in Figure 2.1(b)) suggests that $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ goes to infinity as $B \to \infty$, just as $R_{\boldsymbol{\sigma}}^{\mathrm{sys}}(B)$ (in the centralized approach) does (refer to, Remark 2.1).

This second example gives rise to the main questions that this chapter aims to answer. In particular: what happens to the system sum-rate of NE as the power budget is increased toward infinity? Can we characterize the behaviors observed in the previous numerical example? Consequently, these questions suggest the study of the limit $\lim_{B\to\infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$; which is introduced in Section 2.4 and is studied along this chapter. More specifically, as the power budget goes to infinity, can we compute the finite asymptotic value of the maximum system sum-rate of NE? can we characterize the NE that yields an infinite system sum-rate of NE? Finally, can we draw some conclusions on the efficiency of the noncooperative approach for solving the power allocation problem? Indeed, in the following sections, we provide answers to each of the aforementioned questions.

2.4 Modeling of the System when it is Endowed with Unbounded Power Budgets

In order to answer the questions stated in the previous section we start by describing the general approach that we will follow, this consists basically of the two logical steps outlined next.

- *An auxiliary problem*: We start the analysis with the introduction of a homogenized version of the $\mathrm{MSSR}_{\mathrm{NE}}$ optimization problem defined in (2.9), which we term $\mathrm{HMSSR}_{\mathrm{NE}}$. The latter problem models the limiting behavior of the former one as the budget B goes to infinity. The details of this step are given in Subsection 2.4.1.

- *Relation between the $\mathrm{MSSR}_{\mathrm{NE}}$ and $\mathrm{HMSSR}_{\mathrm{NE}}$ problems*: In Subsection 2.4.2 we present a proposition that summarizes the connections between the $\mathrm{MSSR}_{\mathrm{NE}}$ and its homogenized version. This relation is further explored along this chapter.

For the sake of clarity, in Figure 2.2 we illustrate the progression of our presentation, highlighting the main questions to be answered during the discussion along with the principal results.

2.4.1 An Auxiliary Problem

To construct the $\mathrm{HMSSR}_{\mathrm{NE}}$ problem, we start by rewriting the objective function of the optimization problem (2.9) in such a way that it depends explicitly not only on the power allocation $p_q(k)$ but also on the Lagrange multipliers λ_q. To do so, we begin by using the well-known fact that every feasible solution to (2.9) must satisfy $\lambda_q > 0$ and therefore $B = \sum_{k=1}^{N} p_q(k)$

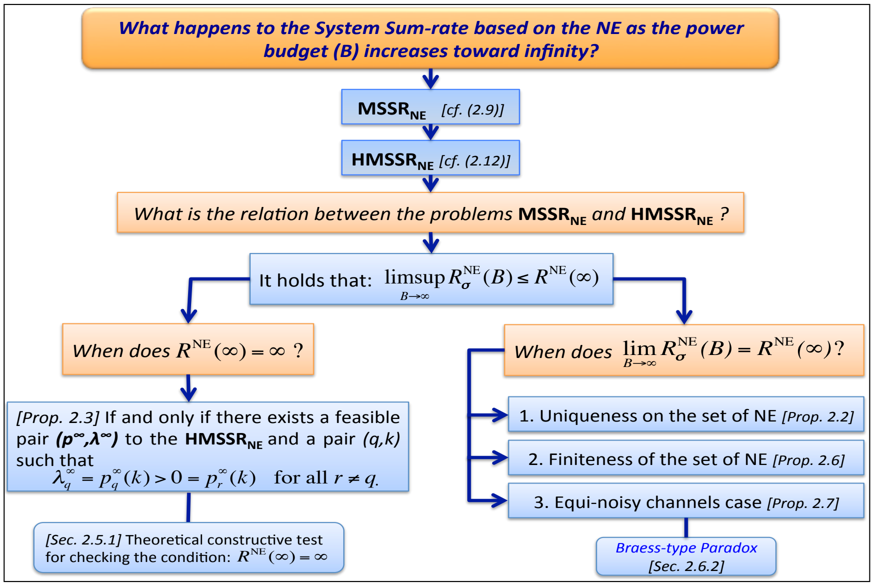

Figure 2.2: General overview of the chapter.

for all $q = 1, \ldots, Q$ i.e. at equilibrium, all users must exhaust their power budget (see, e.g., [68]). Second, we apply the following simple result.

Proposition 2.1. For every pair (q, k), $q = 1, \ldots, Q$ and $k = 1, \ldots, N$ it holds that

$$\log\left(1+\frac{p_q(k)}{\sigma_q^2(k)+\sum_{r\neq q} H_{rq}(k)\,p_r(k)}\right) = \log\lambda_q - \log\left(\lambda_q - p_q(k)\right), \qquad (2.10)$$

for all pairs $(\mathbf{p}, \boldsymbol{\lambda})$ feasible to the problem (2.9).

Proof. Notice that the identity (2.10) obviously holds if $p_q(k) = 0$. If $p_q(k) > 0$, then by complementarity we have:

$$\sigma_q^2(k) + \sum_{r=1}^{Q} H_{rq}(k)\,p_r(k) \;=\; \lambda_q,$$

establishing (2.10) for all (q, k). $\square$

Therefore, using the two observations above, we can write the MSSR$_{\text{NE}}$ in

(2.9) equivalently as:

$$\text{MSSR}_{\text{NE}}: \begin{cases} R_{\boldsymbol{\sigma}}^{\text{NE}}(B) \triangleq \underset{p_q(k),\lambda_q}{\text{maximum}} \displaystyle\sum_{q=1}^{Q}\sum_{k=1}^{N} \left[\log \lambda_q - \log\left(\lambda_q - p_q(k)\right)\right] \\ \text{subject to LCP}(\mathbf{b},\mathbf{M}) \text{ [cf. (2.5)].} \end{cases} \tag{2.11}$$

Finally, we are ready to introduce the homogenization of the LCP($\mathbf{b},\mathbf{M}$) [cf. (2.5)] obtained by setting the power budget of all users equal to unity, and $\sigma_q^2(k) = 0$ for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$. Thus, the HMSSR$_{\text{NE}}$ problem is defined by

$$\text{HMSSR}_{\text{NE}}: \begin{cases} R^{\text{NE}}(\infty) \triangleq \underset{p_q^{\infty}(k),\lambda_q^{\infty}}{\text{maximum}} \displaystyle\sum_{q=1}^{Q}\sum_{k=1}^{N} \left[\log \lambda_q^{\infty} - \log\left(\lambda_q^{\infty} - p_q^{\infty}(k)\right)\right] \\ \text{subject to } \ \forall\, q = 1, \ldots, Q \text{ and } \forall\, k = 1, \ldots, N: \\ \quad 0 \leq \displaystyle\sum_{r=1}^{Q} H_{rq}(k)\, p_r^{\infty}(k) - \lambda_q^{\infty} \quad \perp \quad p_q^{\infty}(k) \geq 0 \\ \quad 0 \leq -1 + \displaystyle\sum_{k=1}^{N} p_q^{\infty}(k) \quad \perp \quad \lambda_q^{\infty} \geq 0. \end{cases} \tag{2.12}$$

Again, the constraints in the HMSSR$_{\text{NE}}$ correspond to a linear complementarity problem LCP($\mathbf{b}_0,\mathbf{M}$), where the matrix $\mathbf{M}$ remains unchanged as in (2.8), and the particular constant vector has the form

$$\mathbf{b}_0 \triangleq \begin{pmatrix} \mathbf{0}_{NQ} \\ -\mathbf{1}_Q \end{pmatrix} \in \mathbb{R}^{\mathbf{N}}, \tag{2.13}$$

where $\mathbf{0}_{NQ}$ is the NQ-dimensional zero vector. It is important to highlight that, intuitively, the LCP($\mathbf{b}_0,\mathbf{M}$) is obtained from the LCP($\mathbf{b},\mathbf{M}$) by normalizing it by the power budget B and taking the limit as $B \to \infty$. As a result, the HMSSR$_{\text{NE}}$ is suitable to study the MSSR$_{\text{NE}}$ in the limit as the power budget increases without bound.

2.4.2 Relation Between the MSSR$_{\text{NE}}$ and HMSSR$_{\text{NE}}$ problems

At this point in the discussion, a natural question to ask is: what is the relation between the MSSR$_{\text{NE}}$ problem in (2.11) and its homogenized version

(HMSSR$_{\text{NE}}$) introduced in (2.12)? The following proposition gives a first insight into the answer of this question.

Proposition 2.2. For any $\boldsymbol{\sigma} > \mathbf{0}$, it holds that $\limsup_{B\to\infty} R_{\boldsymbol{\sigma}}^{\text{NE}}(B) \le R^{\text{NE}}(\infty) \in [\,0,\infty\,]$. Moreover, if (2.12) has a unique feasible solution, then $\lim_{B\to\infty} R_{\boldsymbol{\sigma}}^{\text{NE}}(B) = R^{\text{NE}}(\infty)$.

Proof. Let $\{B^\nu\} \uparrow \infty$ be an arbitrary sequence of unbounded budgets. For each ν, let $(\mathbf{p}^\nu, \boldsymbol{\lambda}^\nu)$, where $\mathbf{p}^\nu \triangleq \left((p_q^\nu(k))_{k=1}^N\right)_{q=1}^Q$ and $\boldsymbol{\lambda}^\nu \triangleq \left(\lambda_q^\nu\right)_{q=1}^Q$, be an optimal pair to (2.11). The normalized sequences $\left\{p_q^\nu(k)/B^\nu\right\}$ and $\left\{\lambda_q^\nu/B^\nu\right\}$ are clearly bounded. By working with an appropriate subsequence, we may assume without loss of generality that, for every pair (q,k),

$$\lim_{\nu\to\infty} \frac{p_q^\nu(k)}{B^\nu} = p_q^\infty(k) \quad \text{and} \quad \lim_{\nu\to\infty} \frac{\lambda_q^\nu}{B^\nu} = \lambda_q^\infty.$$

The pair $(\mathbf{p}^\infty, \boldsymbol{\lambda}^\infty)$, where $\mathbf{p}^\infty \triangleq \left((p_q^\infty(k))_{k=1}^N\right)_{q=1}^Q$ and $\boldsymbol{\lambda}^\infty \triangleq \left(\lambda_q^\infty\right)_{q=1}^Q$, is easily seen to be feasible to (2.12). It follows that $\limsup_{\nu\to\infty} R_{\boldsymbol{\sigma}}^{\text{NE}}(B_\nu) \le R^{\text{NE}}(\infty)$. The second claim of the proposition follows readily from the above proof. $\square$

Notice that Proposition 2.2 relates the optimal objective values of the problems MSSR$_{\text{NE}}$ and HMSSR$_{\text{NE}}$ via the limit (superior) as the power budget B is increased toward infinity. Thus, to characterize the limiting behavior of the maximum system sum-rate of NE i.e. $R_{\boldsymbol{\sigma}}^{\text{NE}}(B)$ as $B \to \infty$, we can simply focus on the optimal objective value of the homogenized problem. Nevertheless, the aforementioned proposition guarantees that the optimal objective values of the MSSR$_{\text{NE}}$ as B goes to infinity coincides with that of the HMSSR$_{\text{NE}}$, whenever the set of NE of the homogenized problem is a singleton. Sufficient conditions ensuring the uniqueness of the NE for the competitive power allocation game have been derived in the literature; see e.g., [68, 97]. Here it is important to highlight that the uniqueness conditions in the cited references readily apply to the homogenized power allocation game (characterized by the LCP$(\mathbf{b}_0, \mathbf{M})$), since they are independent of the noise variances $\boldsymbol{\sigma}$ and the power budget B. Moreover, those conditions are also sufficient for the convergence of the IWFA to the unique NE of the power allocation game.

An immediate consequence of Proposition 2.2 is the following corollary that does not require a proof.

Corollary 2.1. $\lim_{B\to\infty} R^{\text{NE}}_{\boldsymbol{\sigma}}(B) = \infty$ if and only if two conditions hold: (a) $\lim_{B\to\infty} R^{\text{NE}}_{\boldsymbol{\sigma}}(B) = R^{\text{NE}}(\infty)$ and (b) $R^{\text{NE}}(\infty) = \infty$.

In the rest of this chapter, we investigate conditions (a) and (b) separately. In particular, condition (b) is examined in Section 2.5, and in Section 2.6 we provide two more sufficient conditions for (a) to hold.

2.5 Characterizing the Condition $R^{\text{NE}}(\infty) = \infty$

We begin this section by providing the next result that characterizes the condition $R^{\text{NE}}(\infty) = \infty$ in terms of the existence of a NE tuple with a particular form satisfying the constraints of the HMSSR_{NE} problem. This result is based on the fact that, similar to (2.11), for every pair $(\mathbf{p}^{\infty}, \boldsymbol{\lambda}^{\infty})$ feasible to (2.12), we must have $\lambda_q^{\infty} > 0$, and thus $\sum_{k=1}^{N} p_q^{\infty}(k) = 1$ for all $q = 1, \ldots, Q$. As a result, each summand $\left[\log \lambda_q^{\infty} - \log\left(\lambda_q^{\infty} - p_q^{\infty}(k)\right)\right]$ in the objective of $R^{\text{NE}}(\infty)$ either takes on a finite nonnegative value or equals to $+\infty$, the latter occurring if and only if $\lambda_q^{\infty} = p_q^{\infty}(k) > 0$ for at *least one* pair (q, k). This observation yields the following proposition which provides a necessary and sufficient condition for the system sum-rate over the set of NE to reach infinity in the HMSSR_{NE} problem.

Proposition 2.3. $R^{\text{NE}}(\infty) = \infty$ if and only if there exists a pair $(\mathbf{p}^{\infty}, \boldsymbol{\lambda}^{\infty})$ feasible to (2.12) with at least one pair (q, k) satisfying

$$\lambda_q^{\infty} = p_q^{\infty}(k) > 0 = p_r^{\infty}(k) \quad \text{ for all } r \neq q.$$

Proof. The "if" part is obvious. For the "only if" statement, suppose that $R^{\text{NE}}(\infty) = \infty$, thus, there must exist a feasible (hence optimal) pair $(\mathbf{p}^{\infty}, \boldsymbol{\lambda}^{\infty})$ to (2.12) and a pair (q, k) such that $\lambda_q^{\infty} = p_q^{\infty}(k) > 0$. Consequently, by complementarity

$$0 = \sum_{r \neq q} H_{rq}(k)\, p_r^{\infty}(k) \;\;\Rightarrow\;\; p_r^{\infty}(k) = 0 \text{ for all } r \neq q.$$

□

Proposition 2.3 implies that if $\lim_{B\to\infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) = R^{\mathrm{NE}}(\infty)$, for example, when (2.12) has a unique feasible solution as stated in Proposition 2.2, equivalently, when the homogenized game has a unique NE (refer to Section 2.6 for more sufficient conditions). Then, when the power budget in the system increases toward infinity, the maximum system sum-rate of NE will also approach infinity if and only if there exists a user-subchannel pair such that this user allocates power "exclusively" over the corresponding subchannel (i.e. an interference free subchannel) in the homogenized game, which is independent of the tuple $\boldsymbol{\sigma}$. It is important to mention that, even though the case of unbounded budgets is not directly considered in these references, the authors of [92] and [73] observed that higher system sum-rates at the NE are achieved when the users are limited to allocate power in an "exclusive" subchannel. This was observed for the case of two users and two subchannels. Notice that this agrees with our general result given in Proposition 2.3, for the case of unbounded power budgets.

Remark 2.4 (On the flat-fading channels case)**.** A case that is worth exploring is that of flat-fading channels [25] i.e. $H_{rq}(k) = H_{rq}$ for all k and all $r \neq q$, and white noise. When the flat-channel coefficients H_{rq} are sufficiently small, the uniform power allocation (that is, $p_q(k) = p_q(k')$ for all k and k' and all q) happens to be the *unique* Nash equilibrium of the homogenized game (see, [25, Thm. 4]). Then, according to Props. 2.2 and 2.3 we must have $\lim_{B\to\infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) = R^{\mathrm{NE}}(\infty) < \infty$, i.e. the maximum attainable system sum-rate of NE is finite even in the presence of unbounded power budgets. Furthermore, this implies that we have an *asymptotic infinite* price of stability, i.e. $\lim_{B\to\infty} \mathrm{PoS}(B) = \infty$ where $\mathrm{PoS}(B) \triangleq \frac{R_{\boldsymbol{\sigma}}^{\mathrm{sys}}(B)}{R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)}$. In other words, as the power budget increases toward infinity, the noncooperative solution is "infinitely" worse than the one obtained from the cooperative approach. Note that in this case, the price of stability and the price of anarchy coincide (i.e. $\mathrm{PoS}(B) = \mathrm{PoA}(B)$) due to the uniqueness of the NE. This shows the inefficiency of the NE for the particular case of flat-fading channels even when the system's power budget is unbounded. This result is in accordance with [25] that observed the previously described phenomenon in a numerical example for the simpler case of two users (refer to Section 2.3 - Motivating Example 2.2).

In the remaining of this section, we develop a test for verifying the condition $R^{\rm NE}(\infty) = \infty$ based on the the results of Proposition 2.3.

2.5.1 Checking the Condition $R^{\rm NE}(\infty) = \infty$

This section addresses the question of how to check the condition $R^{\rm NE}(\infty) = \infty$ for any particular instance of the HMSSR_{NE} problem in (2.12). We recall that the HMSSR_{NE} is a MPCC whose solvability is, in general, not a trivial task (refer to Remark 2.3). In the intents of verifying such condition, we develop a (theoretical) constructive test via the solution of LCPs. For each pair (q,k), we define the LCP$(\,\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{\,q,k}\,)$:

$$0 \le \widehat{\mathbf{z}}^{\,q,k} \quad \perp \quad \widehat{\mathbf{M}}^{q,k}\, \widehat{\mathbf{z}}^{q,k} + \widehat{\mathbf{b}}_0^{q,k} \ge 0$$

by carrying out the following 3 steps:

(1) Set $\lambda_q^\infty = p_q^\infty(k) = 1 - \sum_{\ell \ne k} p_q^\infty(\ell)$ and substitute λ_q^∞ into the expression: $\sum_{r=1}^{Q} H_{rq}(\ell)\, p_r^\infty(\ell) - \lambda_q^\infty$ for $\ell \ne k$, obtaining

$$\sum_{r=1}^{Q} H_{rq}(\ell)\, p_r^\infty(\ell) - \lambda_q^\infty = 2p_q^\infty(\ell) + \sum_{\ell' \ne k} p_q^\infty(\ell') + \sum_{r \ne q} H_{rq}(\ell)\, p_r^\infty(\ell) - 1.$$

(2) Set $p_r^\infty(k) = 0$ for all $r \ne q$.

(3) Remove, for all $r \ne q$, the constraints:

$$\begin{aligned} 0 \le \sum_{s=1}^{Q} H_{sr}(k)\, p_s^\infty(k) - \lambda_r^\infty &= H_{qr}(k)\, p_q^\infty(k) - \lambda_r^\infty \\ &= H_{qr}(k) \left(1 - \sum_{\ell \ne k} p_q^\infty(\ell) \right) - \lambda_r^\infty. \end{aligned}$$

The variable $\widehat{\mathbf{z}}^{\,q,k}$ in the LCP$(\,\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{\,q,k}\,)$ has dimension $\widehat{\mathbf{N}} \triangleq \mathbf{N} - (Q+1) = NQ - 1$ and has components:

$$\begin{aligned} &\widehat{\mathbf{p}} \triangleq \left(\widehat{\mathbf{p}}_r \triangleq (p_r^\infty(\ell))_{\ell \ne k} \right)_{r=1}^{Q} \in \mathbb{R}^{Q(N-1)} \quad \text{and} \\ &\widehat{\boldsymbol{\lambda}} \triangleq (\,\lambda_r^\infty\,)_{r \ne q} \in \mathbb{R}^{Q-1}. \end{aligned}$$

The matrix $\widehat{\mathbf{M}}^{q,k}$ and vector $\widehat{\mathbf{b}}_0^{q,k}$ are similarly structured as $\mathbf{M}$ and $\mathbf{b}_0$ [cf. (2.8) and (2.13)] respectively, except for the block corresponding to player q. For the sake of clarity, we illustrate the matrix $\widehat{\mathbf{M}}^{q,k}$ and the vector $\widehat{\mathbf{b}}^{q,k}$ for the particular case $(q,k) = (1,1)$ as follows:

$$\widehat{\mathbf{M}}^{1,1} \triangleq \left[\begin{array}{cccc|ccc} \mathbf{A} & \mathrm{Diag}(\widehat{\mathbf{H}}_{12}^1) & \cdots & \mathrm{Diag}(\widehat{\mathbf{H}}_{1Q}^1) & & & \\ \mathrm{Diag}(\widehat{\mathbf{H}}_{21}^1) & \mathbf{I}_{N-1} & \cdots & \mathrm{Diag}(\widehat{\mathbf{H}}_{2Q}^1) & -\mathbf{1}_{N-1} & & \\ \vdots & \vdots & \ddots & \vdots & & \ddots & \\ \mathrm{Diag}(\widehat{\mathbf{H}}_{Q1}^1) & \mathrm{Diag}(\widehat{\mathbf{H}}_{Q2}^1) & \cdots & \mathbf{I}_{N-1} & & & -\mathbf{1}_{N-1} \\ \hline & \mathbf{1}_{N-1}^T & & & & & \\ & & \ddots & & & & \\ & & & \mathbf{1}_{N-1}^T & & & \end{array}\right]$$

where $\widehat{\mathbf{M}}^{1,1} \in \mathbb{R}^{\widehat{\mathbf{N}}\times\widehat{\mathbf{N}}}$, $\widehat{\mathbf{H}}_{sr}^k \triangleq (H_{rs}(\ell))_{\ell\neq k}$ for $s \neq r$, and $\mathbf{A}$ is the special symmetric positive definite matrix

$$\mathbf{A} \triangleq \mathbf{I}_{N-1} + \mathbf{1}_{N-1}\mathbf{1}_{N-1}^T = \begin{bmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & \cdots & 1 & 2 & 1 \\ 1 & \cdots & 1 & 1 & 2 \end{bmatrix} \in \mathbb{R}^{(N-1)\times(N-1)};$$

and

$$\widehat{\mathbf{b}}_0^{1,1} \triangleq \left(\begin{array}{c} \left.\left\{\begin{array}{c} -\mathbf{1}_{N-1} \\ \mathbf{0}_{N-1} \\ \vdots \\ \mathbf{0}_{N-1} \end{array}\right.\right\} Q \text{ blocks} \\ -\mathbf{1}_{Q-1} \end{array} \right) \in \mathbb{R}^{\widehat{\mathbf{N}}}.$$

The following result connects the LCP$(\mathbf{b}_0, \mathbf{M})$ and the so-constructed LCP$(\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{q,k})$.

Proposition 2.4. For a given user-subchannel pair (q,k), if the tuple $\mathbf{z}$ is a solution of the LCP$(\mathbf{b}_0, \mathbf{M})$ such that $p_r^\infty(k) = 0$ for all $r \neq q$, then

(a) $\lambda_r^\infty \leq H_{qr}(k)\left(1 - \sum_{\ell\neq k} p_q^\infty(\ell)\right)$ for all $r \neq q$; and

(b) the tuple $\widehat{\mathbf{z}}^{q,k}$ is a solution of the LCP$(\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{q,k})$.

Conversely, if the tuple $\widehat{\mathbf{z}}^{q,k}$ is a solution of the LCP$(\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{q,k})$ satisfying condition (a), then by letting $p_r^\infty(k) \triangleq 0$ for all $r \neq q$ and $p_q^\infty(k) = \lambda_q^\infty \triangleq 1 - \sum_{\ell \neq k} p_q^\infty(\ell)$, the tuple $\mathbf{z}$ is a solution of the LCP$(\mathbf{b}_0, \mathbf{M})$.

Proof. Suppose that $\mathbf{z}$ is a solution of the LCP$(\mathbf{b}_0, \mathbf{M})$ such that $p_r^\infty(k) = 0$ for all $r \neq q$. We have, for any $r \neq q$,

$$\lambda_r^\infty \leq \sum_{s=1}^{Q} H_{sr}(k)\, p_s^\infty(k) = H_{qr}(k)\, p_q^\infty(k) = \, H_{qr}(k) \left(1 - \sum_{\ell \neq k} p_q^\infty(\ell)\right),$$

which establishes the first assertion of the proposition. Conversely, suppose that $\widehat{\mathbf{z}}^{q,k}$ is a solution of the LCP$(\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{q,k})$ satisfying condition (a). To establish the second assertion of the proposition, it suffices to show that $\sum_{\ell \neq k} p_q^\infty(\ell) < 1$. Assume the contrary, i.e $\sum_{\ell \neq k} p_q^\infty(\ell) \geq 1$, but this contradicts the complementarity condition

$$0 \leq p_q^\infty(\ell) \quad \perp \quad 2p_q^\infty(\ell) + \sum_{\ell' \neq k} p_q^\infty(\ell') + \sum_{r \neq q} H_{rq}(\ell)\, p_r^\infty(\ell) - 1 \geq 0,$$

which holds for all $\ell \neq k$. $\square$

Finally, combining the results in Props. 2.3 and 2.4, we immediately deduce the following corollary that provides a constructive way to ascertain the condition $R^{\mathrm{NE}}(\infty) = \infty$. Note also that this result leads us to verify the condition $\lim_{B \to \infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) = \infty$, provided that there is no "gap" between this limit and $R^{\mathrm{NE}}(\infty)$ as required by Corollary 2.1.

Corollary 2.2. $R^{\mathrm{NE}}(\infty) = \infty$ if and only if there exists a pair (q, k) and a solution $\widehat{\mathbf{z}}^{q,k}$ of the LCP$(\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{q,k})$ satisfying condition (a) of Proposition 2.4.

Based on the previous result, we summarize the test for checking the condition $R^{\mathrm{NE}}(\infty) = \infty$ as follows.

Theoretical Constructive Test of the Condition $R^{\mathrm{NE}}(\infty) = \infty$. For each pair (q, k), solve the LCP$(\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{q,k})$ by any (distributed) algorithm and check if the obtained solution satisfies condition (a) of Proposition 2.4.

If affirmative, then $R^{\mathrm{NE}}(\infty) = \infty$. If no solution of the LCP$(\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{q,k})$ satisfies condition (a) of Proposition 2.4, then $R^{\mathrm{NE}}(\infty) < \infty$.

Notice that, under conditions for the convergence of the well-known IWFA, the test above can be carried out in a distributed manner, by solving possibly many LCPs of the kind $(\widehat{\mathbf{b}}_0^{q,k}, \widehat{\mathbf{M}}^{q,k})$, each of which corresponds to a certain competitive power allocation game.

To conclude this section, we turn our attention to the particular case of two users. Interestingly, in this simpler case, the condition $R^{\mathrm{NE}}(\infty) < \infty$ can be checked by verifying the feasibility of a homogeneous system of linear inequalities, as opposed to solving LCPs as described above for the general case. The following proposition summarizes this result.

Proposition 2.5. Let $Q = 2$ and N be arbitrary. Consider the following four statements:

(a) $R^{\mathrm{NE}}(\infty) < \infty$;

(b) for all solutions $\mathbf{z}^\infty$ of the LCP$(\mathbf{b}_0, \mathbf{M})$, $p_q^\infty(k) > 0$ for all $q = 1, 2$ and $k = 1, \ldots, N$;

(c) the LCP$(\mathbf{b}_0, \mathbf{M})$ has a solution with $p_q^\infty(k) > 0$ for all $q = 1, 2$ and $k = 1, \ldots, N$;

(d) the homogenous equation $\mathbf{M}\,\mathbf{z}^\infty + \mathbf{b}_0\,\tau = 0$ has a solution $(\tau, \mathbf{z}^\infty) > 0$.

It holds that (a) $\Leftrightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d), and (d) $\Rightarrow$ (a) if the LCP$(\mathbf{b}_0, \mathbf{M})$ has a unique solution.

Proof. (a) $\Leftrightarrow$ (b). By way of contradiction, suppose (b) is false, then there exists a solution $(\mathbf{p}^\infty, \boldsymbol{\lambda}^\infty)$ to (2.12) and a pair (q, k), say $q = 2$, such that $p_2^\infty(k) = 0$. Therefore $p_1^\infty(k) - \lambda_1^\infty \geq 0$, which implies $p_1^\infty(k) \geq \lambda_1^\infty > 0$. Hence, by complementarity, $p_1^\infty(k) = \lambda_1^\infty$. By Proposition 2.3, it follows that $R^{\mathrm{NE}}(\infty) = \infty$. Thus (a) $\Rightarrow$ (b). Conversely, if (b) holds, then (a) holds by Proposition 2.3.

(b) $\Rightarrow$ (c). This is clear.

(c) $\Rightarrow$ (d). Under (c), it holds that the equation $\mathbf{M}\mathbf{z}^\infty + \mathbf{b} = 0$ has a solution with $\mathbf{z}^\infty > 0$.

(d) $\Rightarrow$ (a). This is also clear under the uniqueness assumption. $\square$

Contrasting with Proposition 2.3, the significance of this last proposition lies in part (d) which offers a very simple way to verify the condition $R^{\mathrm{NE}}(\infty) < \infty$ (via linear programming) under a uniqueness condition.

2.6 Two Special Cases

Proposition 2.2 provides a sufficient condition for $\lim_{B\to\infty} R^{\mathrm{NE}}_{\boldsymbol{\sigma}}(B) = R^{\mathrm{NE}}(\infty)$, which corresponds to the uniqueness of the NE. In this section, we derive two more sufficient conditions under which equality holds between the optimal objective value of the $\mathrm{HMSSR}_{\mathrm{NE}}$ problem and that of the $\mathrm{MSSR}_{\mathrm{NE}}$ as the budget B grows toward infinity.

2.6.1 The Case of Finitely Many Nash Equilibria

Generalizing the case of a unique NE, we provide another sufficient condition for $\lim_{B\to\infty} R^{\mathrm{NE}}_{\boldsymbol{\sigma}}(B) = R^{\mathrm{NE}}(\infty)$; this condition also implies that the LCP$(\mathbf{b}, \mathbf{M})$ [cf. (2.5)] admits only *finitely many* feasible solutions for all model inputs. Such finiteness of the NE set coupled with a strict complementarity assumption of the solutions of the homogenized LCP$(\mathbf{b}_0, \mathbf{M})$ [cf. (2.8) and (2.13)] provides the desired result. Before formally stating the main result, we introduce some useful definitions and notation.

- We recall that a solution $\mathbf{z}$ of the LCP (2.6) is *strictly complementary* if $\mathbf{z} + \mathbf{b} + \mathbf{M}\mathbf{z} > 0$.

- A real square matrix is said to be *nondegenerate* if all its principal submatrices are nonsingular. Notice that, the matrix $\mathbf{M}$ in (2.8) is not nondegenerate because the lower right block (which is a principal submatrix of $\mathbf{M}$) is composed of only zero elements.

- $\mathbf{A}(k) \in \mathbb{R}^{Q\times Q}$ denotes the matrix of channel gains organized by subchannels i.e. $\mathbf{A}(k) \triangleq [H_{qr}(k)]_{q,r=1}^{Q}$ for all $k = 1, \ldots, N$.

- $\mathbf{A}(k)_{\alpha\alpha} \in \mathbb{R}^{|\alpha|\times|\alpha|}$ denotes the principal submatrix of $\mathbf{A}(k)$ indexed by α, where $\alpha \subseteq \{1, \ldots, Q\}$.

- $\left(\widehat{\mathbf{A}}(k)\right)_{\alpha\alpha} \in \mathbb{R}^{|\alpha|\times|\alpha|}$ denotes the matrix inverse of $\mathbf{A}(k)_{\alpha\alpha}$, whenever the latter is invertible.

- $\mathcal{A}(k;\alpha) \in \mathbb{R}^{Q\times Q}$ denotes the matrix obtained by expanding the matrix $\left(\widehat{\mathbf{A}}(k)\right)_{\alpha\alpha}$ of order $|\alpha|$ into a $Q\times Q$ matrix by filling in with zeros those entries whose indices are not in α.

- By convention, $\mathbf{A}(k)_{\alpha\alpha}$ is the null matrix if $\alpha = \emptyset$, and the corresponding $\mathcal{A}(k;\alpha)$ is the zero matrix.

Taking in consideration the concepts outlined above, we are now ready to present the main result of this subsection.

Proposition 2.6. Under the following two assumptions:

(a) every solution of the LCP$(\mathbf{b}_0, \mathbf{M})$ [cf. (2.8) and (2.13)] is strictly complementary; and,

(b) for every $k = 1, \ldots, N$ the matrix $\mathbf{A}(k)$ is nondegenerate and the matrix $\sum_{k=1}^{N} \mathcal{A}(k;\alpha_k)$ is nonsingular for any $\alpha_k \subseteq \{1,\ldots,Q\}$ such that $\bigcup_{k=1}^{N} \alpha_k = \{1,\ldots,Q\}$,

it holds that $\lim_{B\to\infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) = R^{\mathrm{NE}}(\infty)$. Moreover, condition (b) implies that the problem in (2.11) has only finitely many feasible solutions for all tuples $\boldsymbol{\sigma}$ and power budget B.

Proof. For the sake of clarity, we divide the proof in two parts.

- First, we show that under condition (b) the LCP$(\mathbf{b}, \mathbf{M})$ [cf. (2.5)] has only finitely many feasible solutions. For this purpose, we consider a principal rearrangement of the matrix $\mathbf{M}$ and we rewrite it in the form

$$\mathbf{M} = \left[\begin{array}{cccc|c} \mathbf{A}(1) & & & & -\mathbf{I}_Q \\ & \mathbf{A}(2) & & & -\mathbf{I}_Q \\ & & \ddots & & \vdots \\ & & & \mathbf{A}(N) & -\mathbf{I}_Q \\ \hline \mathbf{I}_Q & \mathbf{I}_Q & \cdots & \mathbf{I}_Q & \end{array}\right],$$

where $\mathbf{I}_Q$ denotes the identity matrix of order Q. We are interested in principal submatrices of $\mathbf{M}$, denoted by $\mathbf{S}^{(0)}$, that are structured as

$$\mathbf{S}^{(0)} \triangleq \left[\begin{array}{c|ccc|c} \mathbf{A}(1)_{\alpha_1\alpha_1} & & & & -\left(\mathbf{I}_Q\right)_{\alpha_1\bullet} \\ \hline & \mathbf{A}(2)_{\alpha_2\alpha_2} & & & -\left(\mathbf{I}_Q\right)_{\alpha_2\bullet} \\ & & \ddots & & \vdots \\ & & & \mathbf{A}(N)_{\alpha_N\alpha_N} & -\left(\mathbf{I}_Q\right)_{\alpha_N\bullet} \\ \hline \left(\mathbf{I}_Q\right)_{\bullet\alpha_1} & \left(\mathbf{I}_Q\right)_{\bullet\alpha_2} & \cdots & \left(\mathbf{I}_Q\right)_{\bullet\alpha_N} & \end{array}\right],$$

where for each $k = 1, \ldots, N$, α_k is a subset of $\{1, \ldots, Q\}$ such that $\bigcup_{k=1}^{N} \alpha_k = \{1, \ldots, Q\}$, and where $\left(\mathbf{I}_Q\right)_{\bullet\alpha_k}$ and $\left(\mathbf{I}_Q\right)_{\alpha_k\bullet}$ are the submatrices of the identity matrix $\mathbf{I}_Q$ consisting of columns and rows indexed by α_k, respectively. We may write the matrix $\mathbf{S}^{(0)}$ as

$$\mathbf{S}^{(0)} = \begin{bmatrix} \mathbf{A}(1)_{\alpha_1\alpha_1} & \mathbf{S}_{12}^{(0)} \\ \mathbf{S}_{21}^{(0)} & \mathbf{S}_{22}^{(0)} \end{bmatrix}$$

where

$$\mathbf{S}_{12}^{(0)} \triangleq \begin{bmatrix} & -\left(\mathbf{I}_Q\right)_{\alpha_1\bullet} \end{bmatrix},$$

$$\mathbf{S}_{21}^{(0)} \triangleq \begin{bmatrix} \\ \\ \left(\mathbf{I}_Q\right)_{\bullet\alpha_1} \end{bmatrix},$$

$$\mathbf{S}_{22}^{(0)} \triangleq \left[\begin{array}{ccc|c} \mathbf{A}(2)_{\alpha_2\alpha_2} & & & -\left(\mathbf{I}_Q\right)_{\alpha_2\bullet} \\ & \ddots & & \vdots \\ & & \mathbf{A}(N)_{\alpha_N\alpha_N} & -\left(\mathbf{I}_Q\right)_{\alpha_N\bullet} \\ \hline \left(\mathbf{I}_Q\right)_{\bullet\alpha_2} & \cdots & \left(\mathbf{I}_Q\right)_{\bullet\alpha_N} & \end{array}\right].$$

Since $\mathbf{S}_{11}^{(0)} \triangleq \mathbf{A}(1)_{\alpha_1\alpha_1}$ is nonsingular by assumption, the Schur-complement of $\mathbf{S}_{11}^{(0)}$ in $\mathbf{S}^{(0)}$ is given by

$$
\begin{aligned}
\mathbf{S}^{(1)} &= \mathbf{S}_{22}^{(0)} - \mathbf{S}_{21}^{(0)} \left(\mathbf{S}_{11}^{(0)}\right)^{-1} \mathbf{S}_{12}^{(0)} \\
&= \left[\begin{array}{c|c|c}
\mathbf{A}(2)_{\alpha_2\alpha_2} & & -\left(\mathbf{I}_Q\right)_{\alpha_2\bullet} \\
\hline
& \ddots & \vdots \\
& \mathbf{A}(N)_{\alpha_N\alpha_N} & -\left(\mathbf{I}_Q\right)_{\alpha_N\bullet} \\
\hline
\left(\mathbf{I}_Q\right)_{\bullet\alpha_2} & \cdots \quad \left(\mathbf{I}_Q\right)_{\bullet\alpha_N} & \left(\mathbf{I}_Q\right)_{\bullet\alpha_1} \left(\mathbf{A}(1)_{\alpha_1\alpha_1}\right)^{-1} \left(\mathbf{I}_Q\right)_{\alpha_1\bullet}
\end{array}\right] \\
&= \begin{bmatrix} \mathbf{A}(2)_{\alpha_2\alpha_2} & \mathbf{S}_{12}^{(1)} \\ \mathbf{S}_{21}^{(1)} & \mathbf{S}_{22}^{(1)} \end{bmatrix}.
\end{aligned}
$$

Using the definitions given at the beginning of Subsection 2.6.1, it is clear that the right-bottom element in the matrix above can be rewritten simply as

$$
\begin{aligned}
\mathbf{S}_{22}^{(1)} &= \left(\mathbf{I}_Q\right)_{\bullet\alpha_1} \left(\mathbf{A}(1)_{\alpha_1\alpha_1}\right)^{-1} \left(\mathbf{I}_Q\right)_{\alpha_1\bullet} \\
&= \left(\mathbf{I}_Q\right)_{\bullet\alpha_1} \left(\widehat{\mathbf{A}}(1)\right)_{\alpha_1\alpha_1} \left(\mathbf{I}_Q\right)_{\alpha_1\bullet} \\
&= \mathcal{A}(1;\alpha_1).
\end{aligned}
$$

By the well-known Schur determinantal formula, it follows that $\mathbf{S}^{(0)}$ is nonsingular if and only if $\mathbf{S}^{(1)}$ is nonsingular. Making a similar partition as before and noting that $\mathbf{S}_{11}^{(1)} \triangleq \mathbf{A}(2)_{\alpha_2\alpha_2}$ is nonsingular, we have that the Schur-complement of $\mathbf{S}_{11}^{(1)}$ in $\mathbf{S}^{(1)}$ is given by

$$
\begin{aligned}
\mathbf{S}^{(2)} &= \mathbf{S}_{22}^{(1)} - \mathbf{S}_{21}^{(1)} \left(\mathbf{S}_{11}^{(1)}\right)^{-1} \mathbf{S}_{12}^{(1)} \\
&= \left[\begin{array}{c|c|c}
\mathbf{A}(3)_{\alpha_3\alpha_3} & & -\left(\mathbf{I}_Q\right)_{\alpha_3\bullet} \\
\hline
& \ddots & \vdots \\
& \mathbf{A}(N)_{\alpha_N\alpha_N} & -\left(\mathbf{I}_Q\right)_{\alpha_N\bullet} \\
\hline
\left(\mathbf{I}_Q\right)_{\bullet\alpha_2} & \cdots \quad \left(\mathbf{I}_Q\right)_{\bullet\alpha_N} & \mathcal{A}(1;\alpha_1) + \mathcal{A}(2;\alpha_2)
\end{array}\right].
\end{aligned}
$$

By the same argument, $\mathbf{S}^{(1)}$ is nonsingular if and only if $\mathbf{S}^{(2)}$ is nonsingular. Continuing this block pivoting procedure up to the N-th block, we deduce that $\mathbf{S}^{(0)}$ is nonsingular if and only if the following

Schur-complement

$$\begin{aligned}\mathbf{S}^{(N)} &= \mathbf{S}_{22}^{(N-1)} - \mathbf{S}_{21}^{(N-1)} \left(\mathbf{S}_{11}^{(N-1)}\right)^{-1} \mathbf{S}_{12}^{(N-1)} \\ &= \sum_{k=1}^{N} \mathcal{A}(k;\alpha_k)\end{aligned}$$

is nonsingular, which is so by assumption. Therefore, any submatrix of $\mathbf{M}$ of the type $\mathbf{S}^{(0)}$ is nonsingular.

For any solution $\mathbf{z}$ of the LCP$(\mathbf{b}, \mathbf{M})$, there must exist a principal submatrix of $\mathbf{M}$ of the type $\mathbf{S}^{(0)}$ such that the positive components of $\mathbf{z}$ must satisfy a square system of linear equations defined by $\mathbf{S}^{(0)}$. Since the latter matrix is nonsingular and there are only finitely many principal submatrices of $\mathbf{M}$ of this type, it follows that the LCP$(\mathbf{b}, \mathbf{M})$ has only finitely many solutions. Purely a consequence of assumption (b), this finiteness result does not depend on the vector $\mathbf{b}$.

- Second, we show that under conditions (a) and (b) it must hold that $R^{\text{NE}}(\infty) = \lim_{B\to\infty} R_{\boldsymbol{\sigma}}^{\text{NE}}(B)$. The proof is based on the construction of a solution for the LCP$(\mathbf{b}, \mathbf{M})$ from any optimal solution of the HMSSR$_{\text{NE}}$ problem (whose objective value is possibly equal to ∞). Let $\mathbf{z}^{\infty} \triangleq \begin{pmatrix} \mathbf{p}^{\infty} \\ \boldsymbol{\lambda}^{\infty} \end{pmatrix} \in \mathbb{R}^{\mathbf{N}}$ be any such optimal solution. Let $\beta \triangleq \{(q,k) \mid p_q^{\infty}(k) > 0\}$ be the index set corresponding to the strictly positive power allocations in $\mathbf{z}^{\infty}$. Let $\bar{\beta}$ be the complement of β in $\{1, \cdots, \mathbf{N}\}$. Using this pair of complementarity index sets β and $\bar{\beta}$, we construct the vector $\mathbf{z} \triangleq \begin{pmatrix} \mathbf{p} \\ \boldsymbol{\lambda} \end{pmatrix} \in \mathbb{R}^{\mathbf{N}}$ as the unique solution to the following system of equations:

$$\begin{aligned} \sigma_q^2(k) + \sum_{r=1}^{Q} H_{rq}(k)\, p_r(k) - \lambda_q \;&=\; 0, \quad \forall\,(q,k) \in \beta \\ \sum_{k=1}^{N} p_q(k) \;&=\; 0, \quad \forall\, q = 1,\ldots,Q \\ p_q(k) \;&=\; 0, \quad \forall\,(q,k) \notin \beta, \end{aligned} \tag{2.14}$$

whose unique solvability is ensured by the nonsingularity of the principal submatrix $\mathbf{S}^{(0)}$. By the nondegeneracy assumption (a), we have $\mathbf{z}^{\infty}_{\beta} > 0$ and $(\mathbf{b} + \mathbf{M}\mathbf{z}^{\infty})_{\bar{\beta}} > 0$. Hence, a sufficiently large $\overline{B} > 0$ can be chosen such that for all $B \geq \overline{B}$,

$$(\,\mathbf{z} + B\,\mathbf{z}^{\infty}\,)_{\beta} \,\geq\, 0 \quad \text{ and } \quad (\,\mathbf{b} + \mathbf{M}\,\mathbf{z} + B\mathbf{M}\,\mathbf{z}^{\infty}\,)_{\bar{\beta}} \,\geq\, 0.$$

Finally, it is easy to see that $\mathbf{z} + B\,\mathbf{z}^{\infty} \in \mathbb{R}^{\mathbf{N}}$ is a solution of the LCP$(\mathbf{b}, \mathbf{M})$. Therefore we have constructed a vector that is feasible to the optimization problem in (2.11). Hence,

$$\begin{aligned}
R^{\mathrm{NE}}_{\boldsymbol{\sigma}}(B) \,\geq & \sum_{q=1}^{Q}\sum_{k=1}^{N}\, \big[\log\big(\lambda_q + B\lambda_q^{\infty}\big) - \log\big(\big(\lambda_q + B\lambda_q^{\infty}\big) \\
& - \big(p_q(k) + Bp_q^{\infty}(k)\big)\big)\,\big] \\
= & \sum_{q=1}^{Q}\sum_{k=1}^{N}\left[\log\left(\frac{\lambda_q}{B} + \lambda_q^{\infty}\right) - \log\left(\left(\frac{\lambda_q}{B} + \lambda_q^{\infty}\right)\right.\right. \\
& \left.\left. - \left(\frac{p_q(k)}{B} + p_q^{\infty}(k)\right)\right)\right].
\end{aligned}$$

Therefore, $\liminf_{B\to\infty} R^{\mathrm{NE}}_{\boldsymbol{\sigma}}(B) \geq R^{\mathrm{NE}}(\infty)$, establishing that equality holds, by Proposition 2.2.

□

Remark 2.5 (On Proposition 2.6 condition (b))**.** Notice that α_k for every $k = 1, \ldots, N$ represent the set of users allocating power along subchannel k; while $\bigcup_{k=1}^{N}\alpha_k = \{1, \ldots, Q\}$ enforces that every user $q = 1, \ldots, Q$ must allocate power along at least one subchannel, since at equilibrium they exhaust their power budget. Thus, in essence, the key role of condition (b) in Proposition 2.6 is to ensure the solvability of the system of linear equations (2.14) for any index set α_k corresponding to the positive power allocations $p_q^{\infty}(k) > 0$ in an optimal solution of the HMSSR$_{\mathrm{NE}}$ problem. A more specialized condition that guarantees such solvability is the positive definiteness of each tone matrix $\mathbf{A}(k)$, or more generally, the "uniform P-property" of these matrices as expressed by condition (17) in [68, Prop. 2]; in turn, the (more restrictive) latter property yields the uniqueness of the NE and the convergence of the IWFA for computing such equilibria.

Furthermore, condition (b) in Proposition 2.6 has an important implication on the *multiplicity of the NE*. More precisely, under Proposition 2.6(b) the set of NE is guaranteed to be finite. In [92], the authors studied the multiplicity of the NE for the particular case of $Q = 2$ users and $N = 2$ subchannels. Let us turn our attention to that particular case. Interestingly, using the fact that the matrices $\mathbf{A}(k)$ are nondegenerate for $k = 1, 2$ (since they are random matrices drawn from continuous distributions), it is not difficult to show that the proposed matrix conditions in Proposition 2.6(b) simplifies to the nonsingularity of the matrix $\sum_{k=1}^{2} \mathcal{A}(k; \alpha_k)$ with $\alpha_1 = \alpha_2 = \{1, 2\}$. Quite surprisingly, the aforementioned condition reduces to the channel gain conditions previously observed in the proof of [92, Corollary 5], where the set of NE (for the power allocation game with $Q = 2$ and $N = 2$) is shown to be finite. As a result, the conditions in Proposition 2.6(b) are not stronger than those proposed in the literature, at least for this particular case. Indeed, to the best of our knowledge, aside from the cited reference and our result, not much is known about the NE multiplicity of the power allocation game.

The combination of the results in Propositions 2.3 and 2.6 yields the following corollary that gives a lower bound on $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ as B tends to infinity. The further implications of this result are discussed in Remark. 2.6.

Corollary 2.3. Under the assumptions of Proposition 2.6, it holds that $\lim\limits_{B \to \infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) = \infty$ if and only if $\liminf\limits_{B \to \infty} \dfrac{R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)}{\log B} \geq 1$.

Proof. It suffices to show the "only if" part. Assume that $\lim\limits_{B \to \infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) = \infty$, then $R^{\mathrm{NE}}(\infty) = \infty$. Hence, by Proposition 2.3, the problem (2.12) has an optimal solution $\mathbf{z}^{\infty} \triangleq \begin{pmatrix} \mathbf{p}^{\infty} \\ \boldsymbol{\lambda}^{\infty} \end{pmatrix} \in \mathbb{R}^{\mathbf{N}}$ such that for some pair (q, k), we have $\lambda_q^{\infty} = p_q^{\infty}(k) > 0 = p_r^{\infty}(k)$ for all $r \neq q$. Let $\mathbf{z} \triangleq \begin{pmatrix} \mathbf{p} \\ \boldsymbol{\lambda} \end{pmatrix}$ be obtained from the system of linear equations (2.14). By the proof of Proposition 2.6, $\widehat{\mathbf{z}} \triangleq \mathbf{z} + B\,\mathbf{z}^{\infty} \in \mathbb{R}^{\mathbf{N}}$ is a solution of the LCP$(\mathbf{b}, \mathbf{M})$ for all $B > 0$ sufficiently large. For this solution, we have $p_q(k) + Bp_q^{\infty}(k) > 0 = p_r(k) + Bp_r^{\infty}(k)$ for all $r \neq q$. Therefore,

$$\begin{aligned} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) &\geq \sum_{s=1}^{Q} \sum_{k=1}^{N} \log \left(1 + \frac{p_s(k) + Bp_s^{\infty}(k)}{\sigma_s^2(k) + \sum_{r \neq s} H_{rs}(k) \left(p_r(k) + Bp_r^{\infty}(k) \right)} \right) \\ &\geq \log \left(1 + \frac{p_q(k) + Bp_q^{\infty}(k)}{\sigma_q^2(k)} \right) \\ &\geq \log \left(\frac{Bp_q^{\infty}(k)}{\sigma_q^2(k)} \right) = \log B + \log \left(\frac{p_q^{\infty}(k)}{\sigma_q^2(k)} \right), \end{aligned}$$

from which the desired conclusion follows readily. $\square$

Remark 2.6 (On the high-SNR slope)**.** Notice that in Corollary 2.3, the quantity $\liminf_{B \to \infty} \dfrac{R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)}{\log B}$ defines the so-called *high-SNR slope* (in the limit inferior sense), see e.g., [46, 64], i.e. the slope of the maximum system sum-rate of NE curve at high SNR, which quantifies the multiplexing gain of the system. Thus, Corollary 2.3 shows that (under the assumptions of Proposition 2.6) the maximum system sum-rate of NE goes to infinity as $B \to \infty$ if and only if the multiplexing gain of the system is at least one. Furthermore, coupling this observation with Proposition 2.3, we have that: when there exits a user-subchannel pair that operates interference free, then a system's multiplexing gain of at least one is achieved. It is worth mentioning that a similar observation, for the simpler case of two users, is given in [73, Lemma 2]. Notice also that, the result in Corollary 2.3 provides a lower bound on the rate at which $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ goes to infinity. Specifically, under the assumptions of Proposition 2.6, if $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ tends to infinity as $B \to \infty$, then $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ tends to infinity at least as fast as $\log B$ does.

2.6.2 The Case of Equi-Noisy Channels

In this subsection, we consider a more restrictive case where the normalized noise variances satisfy the following property:

$$\sigma_q^2(k) = \sigma_q^2(k') = \sigma_q^2 \quad \forall\, k, k' = 1, \ldots, N \text{ and } \forall\, q = 1, \ldots, Q, \tag{2.15}$$

which we call the case of *equi-noisy channels*. Despite its restrictiveness, the aforementioned case has two important implications. First, we are able to provide another sufficient condition under which $\lim_{B\to\infty} R^{\text{NE}}_{\boldsymbol{\sigma}}(B) = R^{\text{NE}}(\infty)$ for any tuple $\boldsymbol{\sigma}$ satisfying the equi-noisy condition (2.15); this result is stated formally in Proposition 2.7. Second, the presence of the *Braess-type paradox* in the Gaussian IC (see, e.g., [3, 2]) is ruled out for this particular case; as shown in Proposition 2.8. It is worth mentioning that the case of flat-fading channels ($H_{rq}(k) = H_{rq}$ for all k and all $r \neq q$) with white noise is an example of the equi-noisy channels case.

Proposition 2.7. Under the equi-noisy channels condition (2.15), it holds that $\lim_{B\to\infty} R^{\text{NE}}_{\boldsymbol{\sigma}}(B) = R^{\text{NE}}(\infty)$. Moreover, $\lim_{B\to\infty} R^{\text{NE}}_{\boldsymbol{\sigma}}(B) = \infty$ if and only if $\liminf_{B\to\infty} \dfrac{R^{\text{NE}}_{\boldsymbol{\sigma}}(B)}{\log B} \geq 1$.

Proof. Let $(\mathbf{p}^\infty, \boldsymbol{\lambda}^\infty)$ be an optimal pair to (2.12) (with the objective value possibly equal to ∞). Under (2.15), let $\lambda_q(\sigma) \triangleq B\lambda_q^\infty + \sigma_q^2$ for all $q = 1, \ldots, Q$. It is not difficult to show that the pair $\left(B\mathbf{p}^\infty, \boldsymbol{\lambda}(\sigma) = (\lambda_q(\sigma))_{q=1}^{Q}\right)$ is feasible to (2.9). Hence,

$$R^{\text{NE}}_{\boldsymbol{\sigma}}(B) \geq \sum_{q=1}^{Q}\sum_{k=1}^{N} \log\left(1 + \frac{B\,p_q^\infty(k)}{\sigma_q^2 + B\sum_{r\neq q} H_{rq}(k)\,p_r^\infty(k)}\right)$$
$$= \sum_{q=1}^{Q}\sum_{k=1}^{N} \log\left(1 + \frac{p_q^\infty(k)}{\dfrac{\sigma_q^2}{B} + \sum_{r\neq q} H_{rq}(k)\,p_r^\infty(k)}\right).$$

Therefore, $\liminf_{B\to\infty} R^{\text{NE}}_{\boldsymbol{\sigma}}(B) \geq R^{\text{NE}}(\infty)$, establishing that equality holds, by Proposition 2.2. The above inequalities also establish the second conclusion of the proposition, as in the proof of Corollary 2.3. $\square$

Note that Proposition 2.7 has also an important implication on the high-SNR slope, as in Corollary 2.3 (see, Remark 2.6).

Interestingly, more can be said in the case of equi-noisy channels. Let us turn our attention to the well-known *Braess paradox* in the context of the Gaussian IC. It turns out that, for this particular case, the optimal objective value $R^{\text{NE}}_{\boldsymbol{\sigma}}(B)$ is a nondecreasing function of the power budget B. As a result, the presence of the Braess-type paradox is prohibited under condition

(2.15). However, it is important to remark that the optimal solutions of the MSSR_{NE} problem are those NE that *maximize* the system sum-rate, thus the Braess-type paradox is ruled out only for those tuples; refer to Section 2.7 - Example 2.1 for an illustration of this idea. The next proposition summarizes this important result.

Proposition 2.8. Under the equi-noisy channels condition (2.15), for any power budgets B and B', it holds that: $R_{\boldsymbol{\sigma}}^{\text{NE}}(B') \geq R_{\boldsymbol{\sigma}}^{\text{NE}}(B)$ if $B' > B$.

Proof. Let $(\mathbf{p}, \boldsymbol{\lambda})$ and $(\mathbf{p}', \boldsymbol{\lambda}')$ be optimal pairs to (2.9) with power budgets B and B', respectively.

First, we show that $(\widehat{\mathbf{p}}, \widehat{\boldsymbol{\lambda}})$, where $\widehat{\mathbf{p}} \triangleq \dfrac{B'}{B}\,\mathbf{p}$ and $\widehat{\lambda}_q \triangleq \dfrac{B'}{B}\,\lambda_q + \left(1 - \dfrac{B'}{B}\right)\sigma_q^2$ for all $q = 1, \ldots, Q$, is feasible to (2.9) with power budget B'. Clearly $\widehat{\mathbf{p}} \geq \mathbf{0}$; and, for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$ we have that

$$\sigma_q^2 + \sum_{r=1}^{Q} H_{rq}(k)\,\widehat{p}_r(k) - \widehat{\lambda}_q = \frac{B'}{B}\left(\sigma_q^2 + \sum_{r=1}^{Q} H_{rq}(k)\,p_r(k) - \lambda_q\right) \geq 0,$$

where the inequality above and the orthogonality follow immediately by the optimality of $(\mathbf{p}, \boldsymbol{\lambda})$ to (2.9) with budget B. Similarly, for all $q = 1, \ldots, Q$ we have that

$$\widehat{\lambda}_q = \frac{B'}{B}\left(\lambda_q - \sigma_q^2\right) + \sigma_q^2 \geq 0,$$

because $\lambda_q - \sigma_q^2 = \displaystyle\sum_{r=1}^{Q} H_{rq}(k)\,p_r(k) - \lambda_q \geq 0$ for all $q = 1, \ldots, Q$. Finally,

$$-B' + \sum_{k=1}^{N} \widehat{p}_q(k) = -B' + \frac{B'}{B}\sum_{k=1}^{N} p_q(k) = 0$$

where, similarly, the last equality follows from the optimality of $(\mathbf{p}, \boldsymbol{\lambda})$ to (2.9) with budget B. Hence, we have shown that $(\widehat{\mathbf{p}}, \widehat{\boldsymbol{\lambda}})$ is feasible to (2.9) with budget B'. As a result, we have that

$$
\begin{aligned}
R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B') &\geq \sum_{q=1}^{Q}\sum_{k=1}^{N}\log\left(1+\frac{\widehat{p}_q(k)}{\sigma_q^2+\sum_{r\neq q}H_{rq}(k)\,\widehat{p}_r(k)}\right)\\
&\overset{(a)}{=} \sum_{q=1}^{Q}\sum_{k=1}^{N}\log\left(1+\frac{p_q(k)}{\dfrac{B}{B'}\,\sigma_q^2+\sum_{r\neq q}H_{rq}(k)\,p_r(k)}\right)\\
&\overset{(b)}{\geq} \sum_{q=1}^{Q}\sum_{k=1}^{N}\log\left(1+\frac{p_q(k)}{\sigma_q^2+\sum_{r\neq q}H_{rq}(k)\,p_r(k)}\right)\\
&= R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B),
\end{aligned}
$$

where (a) follows by letting $\widehat{\mathbf{p}} = \frac{B'}{B}\,\mathbf{p}$ and (b) is a consequence of $\frac{B'}{B} < 1$. $\square$

2.7 Numerical Examples

In this section, we present some examples validating the theory developed throughout this chapter. More specifically, we revisit Motivating Examples 2.1 and 2.2, so that we can contrast our findings with previous observations developed in the literature. The last example in this section illustrates the likelihood of observing the behavior $R^{\mathrm{NE}}(\infty) < \infty$ under practical settings.

Example 2.1. Referring back to the setting of Motivating Example 2.1, it is clear that this example pertains to the case of *equi-noisy* channels with *non-unique* Nash equilibria. This example shows that the validity of Proposition 2.8 is largely due to the fact that the quantity $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ is defined as the *maximum* of the system sum-rates among all such equilibria. This fact can be observed from Table 2.2, where (in contrast to Table 2.1) an alternative NE for the case $B = 2$ is chosen, which corresponds to an orthogonal power allocation that yields a higher data rate, indeed this corresponds to the maximum system sum-rate among the NE. This affirms the monotonicity

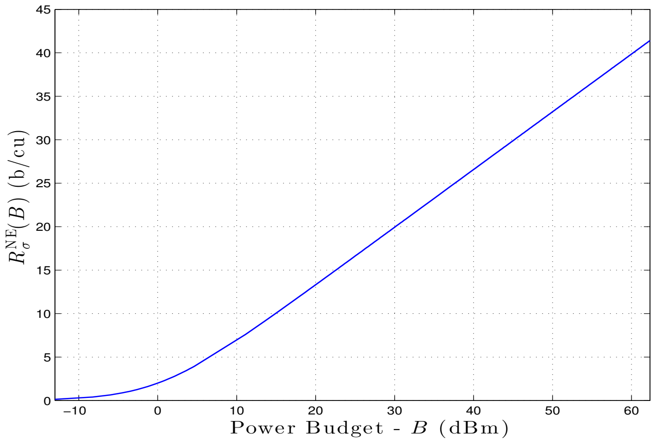

Figure 2.3: Maximum system sum-rate of NE ($R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$) versus power budget ($B$) for the equi-noisy channels example in [67]. As claimed by Proposition 2.8, $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ is a nondecreasing function of B.

of $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$ in the case of equi-noisy channels as claimed by Proposition 2.8. Moreover, by Propositions 2.3 and 2.7 we have that $\lim\limits_{B\to\infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) = \infty$. Figure 2.3 illustrates these observations by showing the nondecreasing behavior of $R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$.

Table 2.2: System sum-rate of NE for the case of non-unique NE. The Braess-type paradox is present for those NE that are not optimal solutions of the $\mathrm{MSSR}_{\mathrm{NE}}$ problem.

Power Budget	NE $(\mathbf{p}_1, \mathbf{p}_2)$	System Sum-Rate of NE	NE $(\mathbf{p}_1^\star, \mathbf{p}_2^\star)$	$R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B)$
$B = 1.5$	$(1.5, 0)$, $(0, 1.5)$	2.64	$(1.5, 0)$, $(0, 1.5)$	2.64
$B = 2$	$(1, 1)$, $(1, 1)$	2.34	$(2, 0)$, $(0, 2)$	3.17

Example 2.2. Consider the system described in Motivating Example 2.2(a), where the authors of [25] pointed out that the distributed system sum-rate approaches $2N\log_2(5)$ as $B \to \infty$. Note that this fact is in accordance with our results since: first, from Proposition 2.2 or invoking the equi-noisy channels case (see, Proposition 2.7), it follows immediately that $\lim\limits_{B\to\infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) = R^{\mathrm{NE}}(\infty)$; and second, by solving the corresponding $\mathrm{HMSSR}_{\mathrm{NE}}$ problem, it is not difficult to obtain that $R^{\mathrm{NE}}(\infty) = 2N\log_2(5)$, as observed in [25]. Clearly, the selfish behavior of the users degrades the performance of the system. To overcome this issue, the authors of the cited paper suggest orthogonal power allocations rather than the uniform ones (obtained from the NE), so that the system sum-rate tends to infinity as $B \to \infty$. This choice is consistent with the results in [37] that relates such a choice of power allocations with the optimality of the maximum centralized sum-rate (2.3).

Example 2.3. In Figure 2.4 we plot the percentage of cases observed where $R^{\mathrm{NE}}(\infty) < \infty$ versus the number of users Q for a fixed number of subchannels $N = 64$. These percentages were calculated by solving the $\mathrm{HMSSR}_{\mathrm{NE}}$ problem (via the IWFA) for 100 different channel realizations per number of users. The channels were simulated as Finite Impulse Response (FIR) filters of order $L = 10$, where the taps are independent and identically distributed zero mean complex Gaussian random variables with variance $1/(d_{rq}^{\gamma}(L+1)^2)$, where d_{rq} denotes the distance between transmitter r and receiver q and γ is the path loss exponent. We assumed $d_{rq} = d_{qr}$ for all $q, r = 1, \ldots .Q$. In these experiments, we only considered those channel realizations satisfying the NE uniqueness condition in [97][Thm. 2], which also guarantees the convergence of the IWFA. The numerical results indicate that, except for the cases of 2, 3 and 4 users, the phenomenon $R^{\mathrm{NE}}(\infty) < \infty$ is observed in 100% of the cases explored. Furthermore, since the NE is unique in all the experiments considered, it happens that $\lim\limits_{B\to\infty} R_{\boldsymbol{\sigma}}^{\mathrm{NE}}(B) < \infty$ holds (for any tuple $\boldsymbol{\sigma}$) as the number of users Q increases. In other words, it does not matter how much we are able to increase the power budget in the system, the maximum attainable sum-rate of NE will always be finite; which implies that we have an asymptotic infinite price of stability/anarchy, i.e. $\lim\limits_{B\to\infty} \mathrm{PoS}(B) = \lim\limits_{B\to\infty} \mathrm{PoA}(B) = \infty$. Consequently, this puts in evidence (at least for the experiments we simulated) the inefficiency of the noncooperative approach for the dynamic power allocation problem, even when the users are endowed with infinite power budgets.

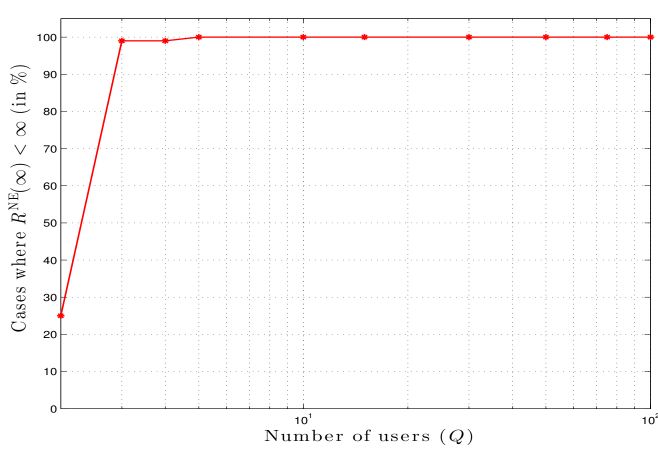

Figure 2.4: Percentage of cases where $R^{\mathrm{NE}}(\infty) < \infty$ versus number of users Q.

2.8 Conclusion

In this chapter, we have considered the problem of maximizing the system sum-rate over the set of NE and analyzed the dependence of this rate on the power budget, in particular when the latter is unbounded. This study touches on the Braess-type paradox and a limiting version of the price of stability. The analysis was primarily based on an auxiliary optimization problem, the $\mathrm{HMSSR}_{\mathrm{NE}}$. We have derived sufficient conditions under which the maximum system sum-rate of NE coincides with the optimal objective value of the homogenized problem in the limit. Such conditions include the uniqueness of the NE, the finiteness of the set of NE, and the special case of equi-noisy channels. Furthermore, for the latter case we have shown that the maximum distributed system sum-rate is a nondecreasing function of the power budget, thus eliminating the presence of the Braess-type paradox. We have further provided a necessary and sufficient condition for the set of NE to reach an infinite system sum-rate when the power budget is unbounded. Moreover, our numerical simulations put in evidence the inefficiency of the noncooperative approach for the power allocation problem, even when the users' power budget is unbounded.

One question that is left open in the present work is a complete characterization for the objective values of the $\mathrm{MSSR}_{\mathrm{NE}}$ and $\mathrm{HMSSR}_{\mathrm{NE}}$ to coincide in the limit. An interesting future research topic is the extension of the present analysis to the case where the users in the communication system are equipped with multiple transmitters and receivers. A major question that this chapter has not addressed is bounding the price of anarchy/stability; this remains a challenging problem and we will continue to study it in our future work.

Chapter 3

A Decomposition Method for Multiuser DC-Programming and its Applications[1]

3.1 Introduction

The resource allocation problem in multiuser systems generally consists of optimizing the (weighted) sum of the users' objective functions, also termed a "social function". In this chapter we address the frequent and difficult case in which the social function is nonconvex and there are (convex) shared constraints coupling the strategies of all the users. Our attention is mainly focused on objective functions of the DC-type, i.e., the difference of two convex functions. It is worth mentioning that DC programs are very common in signal processing, communications, and networking. For instance, the following resource allocation problems belong to the class of DC programs: power control problems in cellular systems [88, 1, 113]; MIMO relay optimization [50]; dynamic spectrum management in DSL systems [117, 111]; sum-rate maximization, proportional-fairness and max-min optimization of SISO/MISO/MIMO ad-hoc networks [51, 93, 41, 94].

In the effort of (optimally) solving DC programs, a great deal of the aforementioned works involves global optimization techniques whose solution methods are mainly based on combinatorial approaches (e.g., adaptations of branch and bound techniques [43]); the results are a variety of *centralized* algorithms customized to the specific DC structure under considerations [88, 117, 1, 113, 50]. However, centralized schemes are too demanding in most applications (e.g., large-scale decentralized networks). This has motivated a number of works whose effort has been finding efficiently high quality (generally locally optimal) solutions of DC programs via easy-to-

[1]This chapter is adapted from [4] ©2014 IEEE. Reprinted, with permision, from A. Alvarado, G. Scutari, and J.-S. Pang, A new decomposition method for multiuser DC-programming and its applications, IEEE Transactions on Signal Processing, June 2014.

implement distributed algorithms. Distributed *ad-hoc* schemes (with provable convergence) for very specific DC formulations have been proposed in [93, 51, 94, 111], mainly based on Successive Convex Approximation (SCA) techniques [94, 90, 5]. In these works however the formulations contain only *private* constraints (i.e., there is no coupling among the users' strategies).

In this work, we move a step forward and consider the more general multiuser DC program, whose feasible set includes also *coupling* convex constraints. To the best of our knowledge, the design of distributed algorithms for this class of problems is an open issue. Indeed, the nonconvexity of the social function prevents the application of standard primal/dual decomposition techniques for convex problems, e.g., [12, 81]; the presence of coupling constraints makes the distributed techniques developed in [93, 51, 94, 111, 90] not directly usable; and standard SCA methods for DC programs [5] wherein the concave part of the objective function is linearized would lead to centralized schemes (because the resulting convex function is generally not separable in the users' variables).

A first contribution of this study is to develop a novel *distributed* decomposition method for solving such a class of multiuser DC problems. Capitalizing on the SCA idea, the proposed novel technique solves a sequence of strongly convex subproblems, whose objective function is obtained by *diagonalization plus off-diagonal linearization of the convex part* of the original DC sum-utility and *linearization of the nonconvex part*. Some desirable features of the proposed approach are: i) Convergence to a stationary solution of the original DC programming is guaranteed also if the subproblems are solved in an inexact way; ii) Each convex subproblem can be distributively solved by the users capitalizing on standard primal or dual decomposition techniques; and iii) It leads to alternative distributed algorithms that differ from rate and robustness of convergence, scalability, local computation versus global communication, and quantity of message passing. All these features make the proposed technique and algorithms applicable to a variety of networks scenarios and problems.

Among the aforementioned applications of the proposed scheme, we show how to customize the developed framework to 2 specific problems, namely:

1) A novel resource allocation problem in the emerging area of cooperative

physical layer security.

2) The renowned sum-rate maximization of MIMO Cognitive Radio (CR) networks.

In summary, the main contributions of this chapter are:

- A novel class of *distributed* decomposition algorithms with provable convergence for multiuser DC problems with (convex) side constraints.

- A novel game theoretical formulation for the secrecy rate maximization problem in *multiple* source-destination OFDMA networks with *multiple* friendly jammers, and consequent algorithms to compute relaxed equilibrium concepts of such game based on its nontrivial DC reformulation.

- A class of *provable convergent* distributed primal/dual algorithms for the CR MIMO sum-rate maximization problem.

The rest of this chapter is organized as follows. Section 3.2 introduces the proposed multiuser DC problem with coupling constraints. Section 3.3 presents a novel decomposition technique for computing stationary solutions of the DC problem, which is suitable for a distributed implementation. Distributed algorithms building on primal and dual decomposition techniques are discussed in Section 3.4. Section 3.5 customizes the proposed framework to the applications in the context of physical layer security and the CR MIMO sum-rate maximization problem. Finally, Section 3.6 draws some conclusions.

3.2 Multiuser DC-Program with Side Constraints

We consider a multiuser system composed of I coupled users. Each user i makes decision on his n_i-dimensional real strategy vector $\mathbf{x}_i \in \mathbb{R}^{n_i}$, subject to some local constraints given by the set $\mathcal{X}_i \subset \mathbb{R}^{n_i}$. The joint strategy set is denoted by $\mathcal{X} \triangleq \prod_{i=1}^{I} \mathcal{X}_i$; $\mathbf{x}_{-i} \triangleq (\mathbf{x}_j)_{j \neq i}$ is the strategy vector of all users except user i; $\mathcal{X}_{-i} \triangleq \prod_{j \neq i} \mathcal{X}_j$; and $\mathbf{x} \triangleq (\mathbf{x}_i)_{i=1}^{I}$ denotes the strategy profile of all the users. In addition to the private constraints, there are also n_c

side constraints of the form $\mathbf{h}(\mathbf{x}) \triangleq (h_j(\mathbf{x}))_{j=1}^{n_c} \leq \mathbf{0}$. The system design is formulated as a DC program in the following form:

$$\begin{aligned}
&\underset{\mathbf{x}}{\text{minimize}} && \theta(\mathbf{x}) \triangleq \sum_{i=1}^{I} (f_i(\mathbf{x}) - g_i(\mathbf{x})) && \\
&\text{subject to} && \mathbf{x}_i \in \mathcal{X}_i \quad \forall\, i = 1, \ldots, I && \text{(private constraints)} \\
& && \mathbf{h}(\mathbf{x}) \leq \mathbf{0} && \text{(coupling constraints).}
\end{aligned} \tag{3.1}$$

Assumptions. We make the following blanket assumptions:

A1) The functions f_i, g_i for $i = 1, \ldots, I$, and h_j for $j = 1, \ldots, n_c$, are convex and continuously differentiable on $\mathcal{X}$.

A2) Each set $\mathcal{X}_i$ is (nonempty) closed and convex.

A3) The functions f_i and g_i have Lipschitz continuous gradients on Ξ with constants $L_{\nabla f_i}$ and $L_{\nabla g_i}$, respectively; where Ξ denotes the convex feasible set of (3.1); let $L_{\nabla\theta} \triangleq \sum_i L_{\nabla f_i} + \sum_i L_{\nabla g_i}$.

A4) The lower level set $\mathcal{L}(\mathbf{x}^0) \triangleq \{\mathbf{x} \in \Xi \,|\, \theta(\mathbf{x}) \leq \theta(\mathbf{x}^0)\}$ of the objective function θ is compact for some $\mathbf{x}^0 \in \Xi$.

A5) The convex coupling constraints are in the separable form: $\mathbf{h}(\mathbf{x}) \triangleq \sum_{i=1}^{I} \mathbf{h}_i(\mathbf{x}_i) \leq \mathbf{0}$.

The assumptions above are quite standard and are satisfied by a large class of practical problems. For instance, A4 guarantees that (3.1) has a solution even when Ξ is not bounded; of course A4 is trivially satisfied if Ξ is bounded.

Instances of problem (3.1) appear in many applications, from signal processing to communications and networking; see Section. 3.5 for some motivating examples. Our goal is to obtain *distributed* best-response-like algorithms for the class of problems (3.1), converging to stationary solutions. This confronts three major challenges, namely: i) the objective function is the sum of differences of two convex functions, and thus in general nonconvex; ii) the objective function is not separable in the users' strategies (each function f_i and g_i depends on the strategy profile $\mathbf{x}$ of all users); and iii) there are side constraints coupling all the optimization variables. We deal with these issues in the following sections.

3.3 A New Best-Response SCA Decomposition

The standard SCA-based technique for DC programs applied to (3.1) would suggest solving a sequence of convex subproblems whose objective function is obtained by linearizing at the current iterate the nonconvex part of $\theta(\mathbf{x})$, that is $-\sum_i g_i(\mathbf{x})$, while retaining the convex part $\sum_i f_i(\mathbf{x})$; see, e.g., [5, 90]. However, notice that the resulting convexified function is not separable in the users' variables $\mathbf{x}_i$ (each f_i depends on the strategy vector of all users $\mathbf{x}$); therefore such SCA techniques will lead to centralized solution methods.

Here we introduce a new decomposition technique that does not suffer from this drawback. To formally describe our approach, let us start rewriting each function $f_i(\mathbf{x})$ as: given $\mathbf{x}^\nu \triangleq (\mathbf{x}_i^\nu)_{i=1}^I \in \Xi$ and denoting $\mathbf{x}_{-i}^\nu \triangleq \left(\mathbf{x}_j^\nu\right)_{j\neq i}$, we have

$$f_i(\mathbf{x}) = f_i(\mathbf{x}_i, \mathbf{x}_{-i}^\nu) + \left[f_i(\mathbf{x}_i, \mathbf{x}_{-i}) - f_i(\mathbf{x}_i, \mathbf{x}_{-i}^\nu)\right]. \tag{3.2}$$

We now approximate the term in brackets using a first order Taylor expansion at $\mathbf{x}^\nu$,

$$f_i(\mathbf{x}_i, \mathbf{x}_{-i}) - f_i(\mathbf{x}_i, \mathbf{x}_{-i}^\nu) \approx \sum_{j\neq i} \nabla_{\mathbf{x}_j} f_i(\mathbf{x}^\nu)^T(\mathbf{x}_j - \mathbf{x}_j^\nu),$$

and approximate (3.2) as

$$f_i(\mathbf{x}) \approx \widetilde{f}_i(\mathbf{x}; \mathbf{x}^\nu) \triangleq \underbrace{f_i(\mathbf{x}_i, \mathbf{x}_{-i}^\nu)}_{\text{diagonalization}} + \underbrace{\sum_{j\neq i} \nabla_{\mathbf{x}_j} f_i(\mathbf{x}^\nu)^T(\mathbf{x}_j - \mathbf{x}_j^\nu)}_{\text{off-diagonal linearization}} \tag{3.3}$$

yielding a *diagonalization* plus *off-diagonal linearization* of $f_i(\mathbf{x})$ at $\mathbf{x}^\nu$. To deal with the nonconvexity of $-g_i(\mathbf{x})$, we replace the functions $g_i(\mathbf{x})$ with its linearization at $\mathbf{x}^\nu$:

$$g_i(\mathbf{x}) \approx \widetilde{g}_i(\mathbf{x}; \mathbf{x}^\nu) \triangleq \underbrace{g_i(\mathbf{x}^\nu) + \sum_{j=1}^{I} \nabla_{\mathbf{x}_j} g_i(\mathbf{x}^\nu)^T(\mathbf{x}_j - \mathbf{x}_j^\nu)}_{\text{linearization}}. \tag{3.4}$$

Based on (3.3) and (3.4), the candidate approximation of the nonconvex sum-utility $\theta(\mathbf{x})$ at $\mathbf{x}^\nu$ is:

$$\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu) \triangleq \sum_{i=1}^{I}\left(\widetilde{f}_i(\mathbf{x};\mathbf{x}^\nu) - \widetilde{g}_i(\mathbf{x};\mathbf{x}^\nu)\right) + \sum_{i=1}^{I}\frac{\tau_i}{2}\|\mathbf{x}_i - \mathbf{x}_i^\nu\|^2, \tag{3.5}$$

where we added a proximal-like regularization term with $\tau_i \geq 0$, whose numerical benefits are well-understood; see, e.g., [12]. Rearranging the terms in the above sum, it is not difficult to see that (3.5) can be equivalently rewritten as

$$\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu) = \sum_{i=1}^{I}\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{x}^\nu), \tag{3.6}$$

where

$$\begin{aligned}\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{x}^\nu) \triangleq &\left[f_i(\mathbf{x}_i,\mathbf{x}_{-i}^\nu) + \sum_{j\neq i}\nabla_{\mathbf{x}_i} f_j(\mathbf{x}^\nu)^T(\mathbf{x}_i - \mathbf{x}_i^\nu)\right] \\ &- \left[g_i(\mathbf{x}^\nu) + \sum_{j=1}^{I}\nabla_{\mathbf{x}_i} g_j(\mathbf{x}^\nu)^T(\mathbf{x}_i - \mathbf{x}_i^\nu)\right] + \frac{\tau_i}{2}\|\mathbf{x}_i - \mathbf{x}_i^\nu\|^2.\end{aligned}$$

Roughly speaking, the main idea behind the above approximations is to use a proper combination of *diagonalization* and *partial linearization* on the convex functions $f_i(\mathbf{x})$ [cf. (3.3)] together with a linearization of the nonconvex terms $g_i(\mathbf{x})$ [cf. (3.4)]. The diagonalization procedure fixes the nonseparability issue in $\theta(\mathbf{x})$ while preserving the convex part in $\theta(\mathbf{x})$, whereas the linearization of $g_i(\mathbf{x})$ gets rid of the nonconvex part in $\theta(\mathbf{x})$. Indeed, this procedure leads to the approximation function $\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu)$ at $\mathbf{x}^\nu$ that is *separable* in the users' variables $\mathbf{x}_i$ (each $\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{x}^\nu)$ depends only on $\mathbf{x}_i$, given $\mathbf{x}^\nu$) and is *strongly convex* in $\mathbf{x} \in \Xi$.

The proposed SCA decomposition consists then in solving iteratively (possibly with a memory) the following sequence of (strongly) convex optimization problems: given $\mathbf{x}^\nu \in \Xi$,

$$\widehat{\mathbf{x}}\left(\mathbf{x}^\nu\right) \triangleq \underset{\mathbf{x}\in\Xi}{\operatorname{argmin}}\ \widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu). \tag{3.7}$$

Proposition 3.1 summarizes the main properties of the best-response map $\widehat{\mathbf{x}}(\bullet)$ defined above. These properties are the key points in establishing the convergence of the proposed algorithm. Lemma 3.1 below is instrumental for the proof of the aforementioned proposition.

Lemma 3.1. Under A1-A4, $\nabla\widetilde{\theta}(\mathbf{x};\bullet)$ is uniformly Lipschitz on Ξ, that is, for any given $\mathbf{x}\in\Xi$,

$$\left\|\nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{x};\mathbf{y})-\nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{x};\mathbf{z})\right\|\leq L_{\nabla\widetilde{\theta}}\,\|\mathbf{z}-\mathbf{y}\|\quad\forall\,\mathbf{y},\mathbf{z}\in\Xi,\tag{3.8}$$

with $L^2_{\nabla\widetilde{\theta}}\triangleq 4\,(L^2_{\nabla\theta}+2\sum_{i=1}^{I}L^2_{\nabla f_i}+\tau^{\max})$, where $L_{\nabla\theta}$ and $L_{\nabla f_i}$ are defined in assumption A3, and $\tau^{\max}\triangleq\max_i\tau_i^2$.

Proof. Invoking the separability of $\widetilde{\theta}$, we have that

$$\left\|\nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{x};\mathbf{y})-\nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{x};\mathbf{z})\right\|^2=\sum_{i=1}^{I}\left\|\nabla_{\mathbf{x}_i}\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{y})-\nabla_{\mathbf{x}_i}\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{z})\right\|^2.$$

Now, rewriting $\nabla_{\mathbf{x}_i}\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{y})-\nabla_{\mathbf{x}_i}\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{z})$ as

$$\begin{aligned}\nabla_{\mathbf{x}_i}\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{y})-\nabla_{\mathbf{x}_i}\widetilde{\theta}_i(\mathbf{x}_i;\mathbf{z})=&\left(\nabla_{\mathbf{x}_i}\theta(\mathbf{y})-\nabla_{\mathbf{x}_i}\theta(\mathbf{z})\right)+\tau_i\left(\mathbf{z}_i-\mathbf{y}_i\right)\\&+\left(\nabla_{\mathbf{x}_i}f_i(\mathbf{x}_i,\mathbf{y}_{-i})-\nabla_{\mathbf{x}_i}f_i(\mathbf{x}_i,\mathbf{z}_{-i})\right)\\&+\left(\nabla_{\mathbf{x}_i}f_i(\mathbf{z})-\nabla_{\mathbf{x}_i}f_i(\mathbf{y})\right),\end{aligned}$$

where we added and subtracted the terms $\nabla_{\mathbf{x}_i}f_i(\mathbf{y})+\nabla_{\mathbf{x}_i}f_i(\mathbf{z})$. By introducing the following definitions:

$$\begin{aligned}\mathbf{r}&\triangleq(\mathbf{r}_i)_{i=1}^{I}\triangleq\left(\nabla_{\mathbf{x}_i}\theta(\mathbf{y})-\nabla_{\mathbf{x}_i}\theta(\mathbf{z})\right)_{i=1}^{I},\\\mathbf{s}&\triangleq(\mathbf{s}_i)_{i=1}^{I}\triangleq\left(\mathbf{z}_i-\mathbf{y}_i\right)_{i=1}^{I},\\\mathbf{t}&\triangleq(\mathbf{t}_i)_{i=1}^{I}\triangleq\left(\nabla_{\mathbf{x}_i}f_i(\mathbf{x}_i,\mathbf{y}_{-i})-\nabla_{\mathbf{x}_i}f_i(\mathbf{x}_i,\mathbf{z}_{-i})\right)_{i=1}^{I},\\\mathbf{u}&\triangleq(\mathbf{u}_i)_{i=1}^{I}\triangleq\left(\nabla_{\mathbf{x}_i}f_i(\mathbf{z})-\nabla_{\mathbf{x}_i}f_i(\mathbf{y})\right)_{i=1}^{I},\end{aligned}$$

(3.8) can be rewritten as

$$\begin{aligned}\left\|\nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{x};\mathbf{y})-\nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{x};\mathbf{z})\right\|^2&=\sum_{i=1}^{I}\|\mathbf{r}_i+\tau_i\mathbf{s}_i+\mathbf{t}_i+\mathbf{u}_i\|^2\\&\leq 4\sum_{i=1}^{I}\left(\|\mathbf{r}_i\|^2+\tau_i^2\|\mathbf{s}_i\|^2+\|\mathbf{t}_i\|^2+\|\mathbf{u}_i\|^2\right).\end{aligned}\tag{3.9}$$

The desired result follows readily by substituting in (3.9) the upper bounds:

$$
\begin{aligned}
\|\mathbf{r}\|^2 &= \|\nabla_{\mathbf{x}}\theta(\mathbf{y}) - \nabla_{\mathbf{x}}\theta(\mathbf{z})\|^2 \leq L^2_{\nabla\theta} \|\mathbf{y}-\mathbf{z}\|^2, \\
\|\mathbf{s}\|^2 &= \|\mathbf{y}-\mathbf{z}\|^2, \\
\|\mathbf{t}\|^2 &\leq \|\mathbf{y}-\mathbf{z}\|^2 \sum_{i=1}^{I} L^2_{\nabla f_i}, \\
\|\mathbf{u}\|^2 &\leq \|\mathbf{y}-\mathbf{z}\|^2 \sum_{i=1}^{I} L^2_{\nabla f_i}.
\end{aligned}
$$

□

In the following result, we let

$$
c_{\boldsymbol{\tau}} \triangleq \min_{i=1,\ldots,I} \left\{ \tau_i + \inf_{\mathbf{x}_{-i}\in\mathcal{X}_{-i}} c_{f_i}(\mathbf{x}_{-i}) \right\}, \tag{3.10}
$$

where $c_{f_i}(\mathbf{x}_{-i}) \geq 0$ is the largest constant such that

$$
(\mathbf{z}_i - \mathbf{w}_i)^T (\nabla_{\mathbf{x}_i} f_i(\mathbf{z}_i, \mathbf{x}_{-i}) - \nabla_{\mathbf{x}_i} f_i(\mathbf{w}_i, \mathbf{x}_{-i})) \geq c_{f_i}(\mathbf{x}_{-i}) \|\mathbf{z}_i - \mathbf{w}_i\|^2,
$$

for all $\mathbf{z}_i, \mathbf{w}_i \in \mathcal{X}_i$ and $\mathbf{x}_{-i} \in \mathcal{X}_{-i}$. Note that $c_{f_i}(\mathbf{x}_{-i}) = 0$ if $f_i(\bullet, \mathbf{x}_{-i})$ is convex but not strongly convex.

Proposition 3.1. Given the DC program (3.1) under A1-A4, the map $\Xi \ni \mathbf{y} \mapsto \widehat{\mathbf{x}}(\mathbf{y}) \in \Xi$ defined in (3.7) has the following properties:

(a) For every given $\mathbf{y} \in \Xi$, the vector $\widehat{\mathbf{x}}(\mathbf{y}) - \mathbf{y}$ is a descent direction of the objective function $\theta(\mathbf{x})$ at $\mathbf{y}$:

$$
(\widehat{\mathbf{x}}(\mathbf{y}) - \mathbf{y})^T \nabla_{\mathbf{x}} \theta(\mathbf{y}) \leq -c_{\boldsymbol{\tau}} \|\widehat{\mathbf{x}}(\mathbf{y}) - \mathbf{y}\|^2, \tag{3.11}
$$

with $c_{\boldsymbol{\tau}}$ defined in (3.10).

(b) $\widehat{\mathbf{x}}(\bullet)$ is Lipschitz continuos on Ξ, with constant $L_{\widehat{\mathbf{x}}} \triangleq L_{\nabla\widetilde{\theta}}/c_{\boldsymbol{\tau}}$, where $L_{\nabla\widetilde{\theta}}$ is defined in Lemma 3.1.

(c) The set of fixed points of $\widehat{\mathbf{x}}(\bullet)$ coincides with the set of stationary solutions of the optimization problem (3.1); therefore $\widehat{\mathbf{x}}(\bullet)$ has a fixed point.

Proof. (a) Given $\mathbf{y} \in \Xi$, by definition, $\widehat{\mathbf{x}}(\mathbf{y})$ satisfies the minimum principle associated with (3.7): for all $\mathbf{z} \triangleq (\mathbf{z}_i)_{i=1}^I \in \Xi$

$$\begin{aligned}
&(\mathbf{z} - \widehat{\mathbf{x}}(\mathbf{y}))^T \nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{y}); \mathbf{y}) \geq 0 \\
\sum_{i=1}^{I} (\mathbf{z}_i - \widehat{\mathbf{x}}_i(\mathbf{y}))^T &\left[\nabla_{\mathbf{x}_i} f_i(\widehat{\mathbf{x}}_i(\mathbf{y}), \mathbf{y}_{-i}) + \sum_{j \neq i} \nabla_{\mathbf{x}_i} f_j(\mathbf{y}) \right. \\
&\left. - \sum_{j=1}^{I} \nabla_{\mathbf{x}_i} g_j(\mathbf{y}) + \tau_i \left(\widehat{\mathbf{x}}_i(\mathbf{y}) - \mathbf{y}_i \right) \right] \geq 0.
\end{aligned} \tag{3.12}$$

Letting $\mathbf{z}_i = \mathbf{y}_i$, and, adding and subtracting $\nabla_{\mathbf{x}_i} f_i(\mathbf{y})$ in each term i of the sum in (3.12), we obtain:

$$\begin{aligned}
(\mathbf{y} - \widehat{\mathbf{x}}(\mathbf{y}))^T \nabla_{\mathbf{x}} \theta(\mathbf{y}) \geq & \sum_{i=1}^{I} (\widehat{\mathbf{x}}_i(\mathbf{y}) - \mathbf{y}_i)^T \left(\nabla_{\mathbf{x}_i} f_i(\widehat{\mathbf{x}}_i(\mathbf{y}), \mathbf{y}_{-i}) - \nabla_{\mathbf{x}_i} f_i(\mathbf{y}) \right) \\
& + \sum_{i=1}^{N} \tau_i \|\widehat{\mathbf{x}}_i(\mathbf{y}) - \mathbf{y}_i\|^2 \geq c_{\boldsymbol{\tau}} \, \|\widehat{\mathbf{x}}(\mathbf{y}) - \mathbf{y}\|^2,
\end{aligned} \tag{3.13}$$

where in the last inequality we used the definition of $c_{\boldsymbol{\tau}}$. This completes the proof of (a).

(b) Let $\mathbf{y}, \mathbf{z} \in \Xi$; by the minimum principle, we have

$$\begin{array}{rcll}
(\mathbf{v} - \widehat{\mathbf{x}}(\mathbf{y}))^T \nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{y}); \mathbf{y}) & \geq & 0 & \forall \mathbf{v} \in \Xi \\
(\mathbf{w} - \widehat{\mathbf{x}}(\mathbf{z}))^T \nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{z}); \mathbf{z}) & \geq & 0 & \forall \mathbf{w} \in \Xi.
\end{array}$$

Setting $\mathbf{v} = \widehat{\mathbf{x}}(\mathbf{z})$ and $\mathbf{w} = \widehat{\mathbf{x}}(\mathbf{y})$ and summing the two inequalities above, after some manipulations, we obtain:

$$\begin{aligned}
& (\widehat{\mathbf{x}}(\mathbf{z}) - \widehat{\mathbf{x}}(\mathbf{y}))^T \left(\nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{z}); \mathbf{z}) - \nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{y}); \mathbf{z}) \right) \\
& \leq (\widehat{\mathbf{x}}(\mathbf{y}) - \widehat{\mathbf{x}}(\mathbf{z}))^T \left(\nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{y}); \mathbf{z}) - \nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{y}); \mathbf{y}) \right).
\end{aligned} \tag{3.14}$$

The Lipschitz property of $\widehat{\mathbf{x}}(\bullet)$, as stated in Proposition 3.1(b), comes from (3.14) using the following lower and upper bounds:

$$(\widehat{\mathbf{x}}(\mathbf{z}) - \widehat{\mathbf{x}}(\mathbf{y}))^T \left(\nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{z}); \mathbf{z}) - \nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{y}); \mathbf{y}) \right) \geq c_{\boldsymbol{\tau}} \, \|\widehat{\mathbf{x}}(\mathbf{z}) - \widehat{\mathbf{x}}(\mathbf{y})\|^2 \tag{3.15}$$

and

$$
\begin{aligned}
(\widehat{\mathbf{x}}(\mathbf{y}) - \widehat{\mathbf{x}}(\mathbf{z}))^T \left(\nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{y}); \mathbf{z}) - \nabla_{\mathbf{x}} \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{y}); \mathbf{y}) \right) \\
\leq L_{\nabla \widetilde{\theta}} \|\widehat{\mathbf{x}}(\mathbf{y}) - \widehat{\mathbf{x}}(\mathbf{z})\| \, \|\mathbf{z} - \mathbf{y}\|
\end{aligned}
\tag{3.16}
$$

where (3.15) is a direct consequence of the uniform strong convexity of $\widetilde{\theta}$, whereas (3.16) follows from the Cauchy-Schwartz inequality and Lemma 3.1. Combining (3.14), (3.15), and (3.16), we obtain the desired result.

(c) First, suppose that $\mathbf{x}^*$ is a fixed point of the map $\widehat{\mathbf{x}}(\mathbf{y})$ i.e. $\mathbf{x}^\star = \widehat{\mathbf{x}}(\mathbf{x}^\star)$. By definition, we have: for all $\mathbf{z} \triangleq (\mathbf{z}_i)_{i=1}^I \in \Xi$,

$$
\begin{aligned}
\sum_{i=1}^{I} (\mathbf{z}_i - \widehat{\mathbf{x}}_i(\mathbf{x}^\star))^T \Bigg[\nabla_{\mathbf{x}_i} f_i(\widehat{\mathbf{x}}_i(\mathbf{x}^\star), \mathbf{x}_{-i}^\star) + \sum_{j \neq i} \nabla_{\mathbf{x}_i} f_j(\mathbf{x}^\star) \\
- \sum_{j=1}^{I} \nabla_{\mathbf{x}_i} g_j(\mathbf{x}^\star) + \tau_i \left(\widehat{\mathbf{x}}_i(\mathbf{x}^\star) - \mathbf{x}_i^\star \right) \Bigg] \geq 0.
\end{aligned}
$$

Using in the above equation $\widehat{\mathbf{x}}(\mathbf{x}^\star) = \mathbf{x}^\star$, we obtain

$$
0 \leq \sum_{i=1}^{I} (\mathbf{z}_i - \mathbf{x}_i^\star)^T \left[\sum_{j=1}^{I} \nabla_{\mathbf{x}_i} f_j(\mathbf{x}^\star) - \sum_{j=1}^{I} \nabla_{\mathbf{x}_i} g_j(\mathbf{x}^\star) \right] = (\mathbf{z} - \mathbf{x}^\star)^T \nabla_{\mathbf{x}} \theta(\mathbf{x}^\star),
$$

implying that $\mathbf{x}^\star$ is a stationary solution of (3.1).

Now, suppose that $\mathbf{x}^\star$ is a stationary solution of (3.1). Notice that $\mathbf{x}^\star$ is an optimal solution of $\min_{\mathbf{x} \in \Xi} \widetilde{\theta}(\mathbf{x}; \mathbf{y})$ since it satisfies the minimum principle. Moreover, $\widehat{\mathbf{x}}(\mathbf{x}^\star)$ is the unique optimal solution of $\min_{\mathbf{x} \in \Xi} \widetilde{\theta}(\mathbf{x}; \mathbf{y})$ with $\mathbf{y} = \mathbf{x}^\star$. Therefore, $\mathbf{x}^\star = \widehat{\mathbf{x}}(\mathbf{x}^\star)$, i.e., $\mathbf{x}^\star$ is a fixed point of $\widehat{\mathbf{x}}(\bullet)$. $\square$

The formal description of the proposed SCA technique is given in Algorithm 3.1. Note that in Step 3 of the algorithm we allow a memory in the update of the iterate $\mathbf{x}^\nu$ in the form of a convex combination via $\gamma^\nu \in (0, 1]$ (this guarantees $\mathbf{x}^{\nu+1} \in \Xi$).

The convergence of the algorithm above is studied in the next theorem.

Theorem 3.1. Given the DC program (3.1) under A1-A4, suppose that $\boldsymbol{\tau} \triangleq (\tau_i)_{i=1}^I$ and $\{\gamma^\nu\}$ are chosen so that one of the two following conditions are satisfied:

Algorithm 3.1 : SCA Algorithm for the DC Program (3.1)

Data: $\boldsymbol{\tau} \triangleq (\tau_i)_{i=1}^I \geq \mathbf{0}, \{\gamma^\nu\} > 0$ and $\mathbf{x}^0 \in \Xi$. Set $\nu = 0$.

(`S.1`): If $\mathbf{x}^\nu$ satisfies a termination criterion, STOP.

(`S.2`): Compute $\widehat{\mathbf{x}}\,(\mathbf{x}^\nu)$ [cf. (3.7)].

(`S.3`): Set $\mathbf{x}^{\nu+1} \triangleq \mathbf{x}^\nu + \gamma^\nu\,(\,\widehat{\mathbf{x}}\,(\mathbf{x}^\nu) - \mathbf{x}^\nu\,)$.

(`S.4`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

(a) *Constant step-size rule:*

$$\gamma^\nu = \gamma \in (0, 1] \quad \forall\, \nu \geq 0 \quad \text{and} \quad 2\, c_{\boldsymbol{\tau}} > \gamma L_{\nabla\theta}. \tag{3.17}$$

with $c_{\boldsymbol{\tau}}$ defined in (3.10).

(b) *Diminishing step-size rule:*

$$c_{\boldsymbol{\tau}} > 0,\;\; \gamma^\nu \in (0, 1],\;\; \gamma^\nu \to 0,\;\; \text{and}\;\; \textstyle\sum_{\nu=0}^\infty \gamma^\nu = +\infty. \tag{3.18}$$

Then, the sequence $\{\mathbf{x}^\nu\}$ generated by Algorithm 3.1 converges in a finite number of iterations to a stationary solution of (3.1) or every limit point of the sequence (at least one such point exists) is a stationary solution of (3.1).

Proof. (a) Under assumptions A1-A4, let $\gamma^\nu = \gamma \in (0, 1]$ for every $\nu \geq 0$, and let $2c_{\boldsymbol{\tau}} > \gamma L_{\nabla\theta}$. Combining the iteration defined in Step 3 of Algorithm 3.1 with the Descent Lemma [12, Prop. A.32] we obtain: for every $\nu \geq 0$

$$\begin{aligned}
\theta(\mathbf{x}^{\nu+1}) &\leq \theta(\mathbf{x}^\nu) + \gamma \nabla_{\mathbf{x}} \theta(\mathbf{x}^\nu)^T\,(\widehat{\mathbf{x}}(\mathbf{x}^\nu) - \mathbf{x}^\nu) + \frac{\gamma^2\, L_{\nabla\theta}}{2} \|\widehat{\mathbf{x}}(\mathbf{x}^\nu) - \mathbf{x}^\nu\|^2 \\
&\leq \theta(\mathbf{x}^\nu) - \gamma c_{\boldsymbol{\tau}} \|\widehat{\mathbf{x}}(\mathbf{x}^\nu) - \mathbf{x}^\nu\|^2 + \frac{\gamma^2 L_{\nabla\theta}}{2} \|\widehat{\mathbf{x}}(\mathbf{x}^\nu) - \mathbf{x}^\nu\|^2 \\
&= \theta(\mathbf{x}^\nu) - \gamma \left(c_{\boldsymbol{\tau}} - \frac{\gamma\, L_{\nabla\theta}}{2} \right) \|\widehat{\mathbf{x}}(\mathbf{x}^\nu) - \mathbf{x}^\nu\|^2
\end{aligned}$$

where the second inequality follows from Proposition 3.1(a). Therefore, since by assumption θ is bounded below on Ξ and $c_{\boldsymbol{\tau}} > \frac{\gamma}{2} L_{\nabla\theta}$, then the sequence $\{\theta(\mathbf{x}^\nu)\}$ converges and

$$\lim_{\nu\to\infty} \|\widehat{\mathbf{x}}(\mathbf{x}^\nu) - \mathbf{x}^\nu\| = 0. \tag{3.19}$$

Finally, by assumption A4, the sequence $\{\mathbf{x}^\nu\}$ is bounded, thus it has at least one limit point, let it be $\mathbf{x}^\infty \in \Xi$. Moreover, from Proposition 3.1(b) we know that the map $\widehat{\mathbf{x}}(\bullet)$ is continuous over Ξ, and from (3.19), we must have that $\mathbf{x}^\infty = \widehat{\mathbf{x}}(\mathbf{x}^\infty)$. By Proposition 3.1(c) we have that the set of fixed points of $\widehat{\mathbf{x}}(\bullet)$ coincides with the set of stationary solutions of (3.1), hence $\mathbf{x}^\infty$ is a stationary solution of (3.1).

(b) The proof follows from [94, Th.3] and Proposition 3.1, which establishes the main properties of the best-response map $\Xi \ni \mathbf{y} \mapsto \widehat{\mathbf{x}}(\mathbf{y}) \in \Xi$ defined in (3.7), as required by [94, Th.3]. □

Remark 3.1 (On Algorithm 3.1)**.** The algorithm implements a novel SCA decomposition technique: at each iteration ν, a separable (strongly) convex function $\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu)$ is minimized over the convex set Ξ. The main difference from the classical SCA techniques (e.g., [43, 5, 90]) is that the approximation function $\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu)$ is separable across the users. In Section 3.4, we show that such a structure leads naturally to a distributed implementation of Algorithm 3.1. A practical termination criterion in Step 1 is to stop the iterates when $|\theta(\mathbf{x}^\nu) - \theta(\mathbf{x}^{\nu-1})| \leq \delta$, where δ is a prescribed accuracy. Finally, it is reasonable to expect the algorithm to perform better than classical gradient algorithms applied directly to (3.1) at the cost of no extra signaling, because the structure of the objective function $\theta(\mathbf{x})$ is better explored.

On the choice of the parameters. Theorem 3.1 offers some flexibility in the choice of $\boldsymbol{\tau}$ and γ^ν, while guaranteeing convergence of Algorithm 3.1, which makes it applicable in a variety of scenarios. More specifically, one can use a constant or a diminishing rule for the step-size γ^ν.

Constant step-size rule: This rule resembles analogous constant step-size rules in gradient algorithms: convergence is guaranteed either under "sufficiently" small step-size γ (given $\boldsymbol{\tau}$) or "sufficiently" large τ_i's (given $\gamma \in (0,1]$), such that (3.17) is satisfied. A special case that is worth mentioning is: $\gamma = 1$ and $2\,c_{\boldsymbol{\tau}} > \gamma L_{\nabla\theta}$, which leads to a SCA-based iterate with no memory: given $\mathbf{x}^\nu \in \Xi$, $\mathbf{x}^{\nu+1} \triangleq \widehat{\mathbf{x}}\,(\mathbf{x}^\nu)$. In this particular case, we can relax a bit the convergence condition (3.17) and require the Lipschitz continuity of the gradients of f_i only. The result is stated next.

Theorem 3.2. Let $\{\mathbf{x}^\nu\}$ be the sequence generated by Algorithm 3.1, in the setting of Theorem 3.1 where however we relax A3 by assuming that f_i have Lipschitz gradients with constants $L_{\nabla f_i}$. Suppose that $\gamma^\nu = 1$ for all ν and

$\boldsymbol{\tau} > \mathbf{0}$ is such that $\tau^{\min} \triangleq \min_i \tau_i > 2\sum_{i=1}^{I} L_{\nabla f_i}$; then, the conclusions of Theorem 3.1 hold.

Proof. By induction, since $\mathbf{x}^\nu \in \Xi$, we have

$$\begin{aligned}
\sum_{i=1}^{I}\left[f_i(\mathbf{x}^\nu) - g_i(\mathbf{x}^\nu)\right] \geq \sum_{i=1}^{I}\Bigg[&\Bigg(f_i(\mathbf{x}_i^{\nu+1}, \mathbf{x}_{-i}^\nu) + \sum_{j\neq i}\nabla_{\mathbf{x}_j} f_i(\mathbf{x}^\nu)^T(\mathbf{x}_j^{\nu+1} - \mathbf{x}_j^\nu)\Bigg) \\
&- \Bigg(g_i(\mathbf{x}^\nu) + \sum_{j=1}^{I}\nabla_{\mathbf{x}_j} g_i(\mathbf{x}^\nu)^T(\mathbf{x}_j^{\nu+1} - \mathbf{x}_j^\nu)\Bigg) \\
&+ \frac{\tau_i}{2}\|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^\nu\|^2\Bigg] \\
= \sum_{i=1}^{I}\Bigg[&\Bigg(f_i(\mathbf{x}_i^{\nu+1}, \mathbf{x}_{-i}^\nu) + \sum_{j\neq i}\nabla_{\mathbf{x}_j} f_i(\mathbf{x}^\nu)^T(\mathbf{x}_j^{\nu+1} - \mathbf{x}_j^\nu)\Bigg) \\
&- \left(g_i(\mathbf{x}^\nu) + \nabla_{\mathbf{x}} g_i(\mathbf{x}^\nu)^T(\mathbf{x}^{\nu+1} - \mathbf{x}^\nu)\right) + \frac{\tau_i}{2}\|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^\nu\|^2\Bigg] \\
\geq \sum_{i=1}^{I}\Bigg[&\Bigg(f_i(\mathbf{x}_i^{\nu+1}, \mathbf{x}_{-i}^\nu) + \sum_{j\neq i}\nabla_{\mathbf{x}_j} f_i(\mathbf{x}^\nu)^T(\mathbf{x}_j^{\nu+1} - \mathbf{x}_j^\nu)\Bigg) \\
&- g_i(\mathbf{x}^{\nu+1}) + \frac{\tau_i}{2}\|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^\nu\|^2\Bigg],
\end{aligned}$$

where the last inequality follows by convexity of each function g_i for $i = 1, \ldots, I$. After adding and subtracting the terms $f_i(\mathbf{x}^{\nu+1})$ for $i = 1, \cdots, I$ in the right hand side of the last inequality we obtain

$$\begin{aligned}
\theta(\mathbf{x}^\nu) \geq \theta(\mathbf{x}^{\nu+1}) + \sum_{i=1}^{I}\Bigg[&\left(f_i(\mathbf{x}_i^{\nu+1}, \mathbf{x}_{-i}^\nu) - f_i(\mathbf{x}^{\nu+1})\right) \\
&+ \sum_{j\neq i}\nabla_{\mathbf{x}_j} f_i(\mathbf{x}^\nu)^T(\mathbf{x}_j^{\nu+1} - \mathbf{x}_j^\nu) + \frac{\tau_i}{2}\|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^\nu\|^2\Bigg] \\
= \theta(\mathbf{x}^{\nu+1}) + \sum_{i=1}^{I}\Bigg[&\left(f_i(\mathbf{x}_i^{\nu+1}, \mathbf{x}_{-i}^\nu) - f_i(\mathbf{x}^\nu)\right) \\
&+ \left(f_i(\mathbf{x}^\nu) - f_i(\mathbf{x}^{\nu+1})\right) + \sum_{j\neq i}\nabla_{\mathbf{x}_j} f_i(\mathbf{x}^\nu)^T(\mathbf{x}_j^{\nu+1} - \mathbf{x}_j^\nu) \\
&+ \frac{\tau_i}{2}\|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^\nu\|^2\Bigg].
\end{aligned}$$

By invoking the convexity of each f_i $(i = 1, \ldots, I)$ we have that

$$
\begin{aligned}
f_i(\mathbf{x}_i^{\nu+1}, \mathbf{x}_{-i}^{\nu}) - f_i(\mathbf{x}^{\nu}) &\geq \nabla_{\mathbf{x}_i} f_i(\mathbf{x}^{\nu})^T (\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^{\nu}) \\
\text{and } f_i(\mathbf{x}^{\nu}) - f_i(\mathbf{x}^{\nu+1}) &\geq \nabla_{\mathbf{x}} f_i(\mathbf{x}^{\nu+1})^T (\mathbf{x}^{\nu} - \mathbf{x}^{\nu+1}).
\end{aligned}
$$

As a result,

$$
\begin{aligned}
\theta(\mathbf{x}^{\nu}) \geq{}& \theta(\mathbf{x}^{\nu+1}) + \sum_{i=1}^{I} \Big[\big(\nabla_{\mathbf{x}_i} f_i(\mathbf{x}^{\nu})^T (\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^{\nu}) \big) + \big(\nabla_{\mathbf{x}} f_i(\mathbf{x}^{\nu+1})^T (\mathbf{x}^{\nu} - \mathbf{x}^{\nu+1}) \big) \\
& + \sum_{j \neq i} \nabla_{\mathbf{x}_j} f_i(\mathbf{x}^{\nu})^T (\mathbf{x}_j^{\nu+1} - \mathbf{x}_j^{\nu}) + \frac{\tau_i}{2} \|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^{\nu}\|^2 \Big] \\
={}& \theta(\mathbf{x}^{\nu+1}) - \sum_{i=1}^{I} \Big[\big(\nabla_{\mathbf{x}} f_i(\mathbf{x}^{\nu+1}) - \nabla_{\mathbf{x}} f_i(\mathbf{x}^{\nu}) \big)^T (\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu}) \\
& - \frac{\tau_i}{2} \|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^{\nu}\|^2 \Big] \\
\overset{(a)}{\geq}{}& \theta(\mathbf{x}^{\nu+1}) - \sum_{i=1}^{I} \Big[\|\nabla_{\mathbf{x}} f_i(\mathbf{x}^{\nu+1}) - \nabla_{\mathbf{x}} f_i(\mathbf{x}^{\nu})\| \, \|\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu}\| \\
& - \frac{\tau_i}{2} \|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^{\nu}\|^2 \Big] \\
\overset{(b)}{\geq}{}& \theta(\mathbf{x}^{\nu+1}) - \sum_{i=1}^{I} \Big[L_{\nabla f_i} \|\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu}\|^2 - \frac{\tau_i}{2} \|\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^{\nu}\|^2 \Big] \\
\overset{(c)}{\geq}{}& \theta(\mathbf{x}^{\nu+1}) + \left(\frac{\tau^{\min}}{2} - \sum_{i=1}^{I} L_{\nabla f_i} \right) \|\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu}\|^2
\end{aligned}
$$

where: (a) follows by Cauchy-Schwarz inequality, (b) is a direct consequence of the Lipschitz continuity of the gradients of each function f_i, and (c) follows by taking $\tau^{\min} \triangleq \min_i \tau_i$. Therefore, since by assumption θ is bounded below on Ξ and $\boldsymbol{\tau} > \mathbf{0}$ is chosen such that $\tau^{\min} > 2 \sum_{i=1}^{N} L_{\nabla f_i}$, then the sequence $\{\theta(\mathbf{x}^{\nu})\}$ converges and

$$
\lim_{\nu \to \infty} \|\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu}\| = 0.
$$

By the minimum principle, we have for every $\nu \geq 0$ and for all $\mathbf{x} \in \Xi$

$$
\left(\mathbf{x} - \mathbf{x}^{\nu+1} \right)^T \nabla_{\mathbf{x}} \widetilde{\theta}(\mathbf{x}^{\nu+1}; \mathbf{x}^{\nu}) = \sum_{i=1}^{I} \left(\mathbf{x}_i - \mathbf{x}_i^{\nu+1} \right)^T \nabla_{\mathbf{x}_i} \widetilde{\theta}(\mathbf{x}_i^{\nu+1}; \mathbf{x}^{\nu}) \geq 0 \quad (3.20)
$$

where

$$\nabla_{\mathbf{x}_i}\widetilde{\theta}(\mathbf{x}_i^{\nu+1};\mathbf{x}^\nu) = \nabla_{\mathbf{x}_i} f_i(\mathbf{x}_i^{\nu+1},\mathbf{x}_{-i}^\nu) + \sum_{j\neq i}\nabla_{\mathbf{x}_i} f_j(\mathbf{x}^\nu) - \sum_{j=1}^{I}\nabla_{\mathbf{x}_i} g_j(\mathbf{x}^\nu)$$
$$+ \tau_i\left(\mathbf{x}_i^{\nu+1} - \mathbf{x}_i^\nu\right).$$

Finally, by assumption A4, the sequence $\{\mathbf{x}^\nu\}$ is bounded. Thus, it has at least one limit point, let it be $\mathbf{x}^\infty \triangleq (\mathbf{x}_i^\infty)_{i=1}^I$; then passing the limit $\nu \to \infty$ along an appropriate subsequence (in (3.20)) we get, for all $\mathbf{x} \in \Xi$

$$\begin{aligned}
0 &\leq \sum_{i=1}^{I}\left(\mathbf{x}_i - \mathbf{x}_i^\infty\right)^T\left[\sum_{j=1}^{I}\left(\nabla_{\mathbf{x}_i} f_j(\mathbf{x}^\infty) - \nabla_{\mathbf{x}_i} g_j(\mathbf{x}^\infty)\right)\right] \\
&= \sum_{i=1}^{I}\left(\mathbf{x}_i - \mathbf{x}_i^\infty\right)^T\nabla_{\mathbf{x}_i}\theta(\mathbf{x}^\infty) = \left(\mathbf{x} - \mathbf{x}^\infty\right)^T\nabla\theta(\mathbf{x}^\infty)
\end{aligned}$$

i.e. $\mathbf{x}^\infty$ is a stationary solution of (3.1). □

Diminishing step-size rule: The application of a constant step-size rule requires the knowledge of the Lipschitz constants $L_{\nabla f_i}$ and $L_{\nabla g_i}$, which may not be available. One can use a (conservative) estimate of such values (e.g., using upper bounds), but in practice this generally leads to "large" values of $c_{\boldsymbol{\tau}}$ satisfying (3.17), which reasonably slows down the algorithm. In all these situations, a valid alternative is to use a diminishing step-size in the form (3.18); examples of such rules are [94]: given $\gamma^0 = 1$,

$$\text{Rule 1:} \qquad \gamma^\nu = \gamma^{\nu-1}\left(1 - \epsilon\,\gamma^{\nu-1}\right), \quad \nu = 1,\ldots, \tag{3.21}$$
$$\text{Rule 2:} \qquad \gamma^\nu = \tfrac{\gamma^{\nu-1}+\beta_1}{1+\beta_2\nu}, \qquad \nu = 1,\ldots, \tag{3.22}$$

where $\epsilon \in (0,1)$ and $\beta_1, \beta_2 \in (0,1)$ are given constants such that $\beta_1 \leq \beta_2$.

3.3.1 Inexact implementation of $\widehat{\mathbf{x}}\left(\mathbf{x}^\nu\right)$

We can reduce the computational effort of Algorithm 3.1 by allowing inexact computations of the solution $\widehat{\mathbf{x}}\left(\mathbf{x}^\nu\right)$. The convergence of the resulting algorithm is still guaranteed under more stringent conditions on the step-size and some requirements on the computational errors. The inexact version of Algorithm 3.1 is formally described in Algorithm 3.2 below, where Step 2 of Algorithm 3.1, the exact computation of $\widehat{\mathbf{x}}(\mathbf{x}^\nu)$, is replaced now with its inexact version, that is find a $\mathbf{z}^\nu$ such that $\|\mathbf{z}^\nu - \widehat{\mathbf{x}}(\mathbf{x}^\nu)\| \leq \varepsilon^\nu$, with ε^ν being the accuracy in the computation of $\widehat{\mathbf{x}}(\mathbf{x}^\nu)$ at iteration ν.

Algorithm 3.2: Inexact version of Algorithm 3.1

Data: $\boldsymbol{\tau} \geq \mathbf{0}, \{\gamma^\nu\} > 0, \{\varepsilon^\nu\} \downarrow 0$, and $\mathbf{x}^0 \in \Xi$. Set $\nu = 0$.

(S.1): If $\mathbf{x}^\nu$ satisfies a termination criterion, STOP.

(S.2): Find a $\mathbf{z}^\nu$ such that $\|\mathbf{z}^\nu - \widehat{\mathbf{x}}(\mathbf{x}^\nu)\| \leq \varepsilon^\nu$.

(S.3): Set $\mathbf{x}^{\nu+1} \triangleq \mathbf{x}^\nu + \gamma^\nu(\mathbf{z}^\nu - \mathbf{x}^\nu)$.

(S.4): $\nu \leftarrow \nu + 1$ and go to (S.1).

Convergence is guaranteed if the error sequence $\{\varepsilon^\nu\}$ and the step-size $\{\gamma^\nu\}$ are properly chosen, as stated in Theorem 3.3. The proof of this result is based on the application of Proposition 3.1 and [94, Th. 4], and is omitted.

Theorem 3.3. Let $\{\mathbf{x}^\nu\}$ be the sequence generated by Algorithm 3.2 in the setting of Theorem 3.1, where however A4 is strengthened by assuming that θ is coercive on Ξ, and with the additional assumption that $\nabla_{\mathbf{x}}\theta$ is bounded on Ξ. Suppose that $\{\gamma^\nu\}$ and $\{\varepsilon^\nu\}$ are chosen so that the following conditions are satisfied: i) $\gamma^\nu \in (0,1]$; ii) $\gamma^\nu \to 0$; iii) $\sum_{\nu=0}^\infty \gamma^\nu = +\infty$; iv) $\sum_{\nu=0}^\infty (\gamma^\nu)^2 < +\infty$; and v) $\sum_{\nu=0}^\infty \varepsilon^\nu \gamma^\nu < +\infty$. Then, the conclusions of Theorem 3.1 hold.

Note that the steps-size rule in (3.21) satisfies the square summability condition in Theorem 3.3. As expected, in the presence of errors, convergence of Algorithm 3.2 is guaranteed if $\varepsilon^\nu \to 0$, meaning that the sequence of approximated problems (3.7) is solved with increasing accuracy. Note that Theorem 3.3(v) imposes also a constraint on the rate by which ε^ν goes to zero, which depends also on $\{\gamma^\nu\}$. An example of an error sequence satisfying condition (v) is $\varepsilon^\nu \leq \alpha\,\gamma^\nu$, where α is any finite positive constant. Interestingly, such a condition can be enforced in Algorithm 3.2 using classical error bound results in convex analysis; see, e.g., [28, Ch. 6]. Two examples of error bounds are given in Lemma 3.2 below, where we introduced the following residual quantities:

$$\begin{aligned} r_{\Xi}(\mathbf{z};\mathbf{x}^{\nu}) &\triangleq \left\|\Pi_{\mathcal{N}(\mathbf{z},\Xi)}(-\nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{z};\mathbf{x}^{\nu})) + \nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{z};\mathbf{x}^{\nu})\right\| \\ l_{\Xi}(\mathbf{z};\mathbf{x}^{\nu}) &\triangleq \left\|\mathbf{z} - \Pi_{\Xi}\left(\mathbf{z} - \nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{z};\mathbf{x}^{\nu})\right)\right\| \end{aligned}$$

with $\Pi_{\mathcal{N}(\mathbf{z},\Xi)}(\mathbf{y})$ [resp. $\Pi_{\Xi}(\mathbf{y})$] denoting the Euclidean projection of $\mathbf{y}$ onto the normal cone $\mathcal{N}(\mathbf{z},\Xi) \triangleq \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y}^T(\mathbf{x}-\mathbf{z}) \leq 0, \forall \mathbf{x} \in \Xi\}$ (resp. Ξ).

Lemma 3.2. Given the optimization problem (3.7) under assumptions A1-A4, the following hold:

(a) A (finite) constant $\xi_1 > 0$ exists such that

$$\|\mathbf{z} - \widehat{\mathbf{x}}\,(\mathbf{x}^{\nu})\,\| \leq \xi_1\, r_{\Xi}(\mathbf{z};\mathbf{x}^{\nu}), \quad \mathbf{z} \in \Xi;$$

(b) A (finite) constant $\xi_2 > 0$ exists such that

$$\|\mathbf{z} - \widehat{\mathbf{x}}\,(\mathbf{x}^{\nu})\,\| \leq \xi_2\, l_{\Xi}(\mathbf{z};\mathbf{x}^{\nu}), \quad \mathbf{z} \in \mathbb{R}^n.$$

Proof. This proof is similar to that of [28, Prop. 6.3.1, Prop. 6.3.7]. As a technical note, referring to the cited propositions and using the same notation therein, we emphasize that the proof of the aforementioned propositions can be easily extended to the case of non-Cartesian set K but strongly monotone VI functions $\mathbf{F}$. For the sake of completeness, we state and prove both extensions (see, Propositions 3.2 and 3.3 below). Results (a) and (b) of Lemma 3.2 are direct consequences of applying Propositions 3.2 and 3.3 respectively, to the VI$(\Xi, \nabla_{\mathbf{x}}\widetilde{\theta}(\mathbf{z};\mathbf{x}^{\nu}))$, $\mathbf{z} \in \Xi$. □

Proposition 3.2. Consider the VI$(\mathcal{K},\mathbf{F})$. Let $\mathcal{K} \subset \mathbb{R}^n$ be a closed and convex set. Let the map $\mathbf{F}:\mathbb{R}^n \to \mathbb{R}^n$ be continuous and strongly monotone on $\mathcal{K}$ with constant $\eta > 0$ i.e. $(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y}))^T(\mathbf{x}-\mathbf{y}) \geq \eta\|\mathbf{x}-\mathbf{y}\|^2$ for all $\mathbf{x},\mathbf{y} \in \mathcal{K}$. Let $\mathbf{x}^{\star}$ be the unique solution of the VI$(\mathcal{K},\mathbf{F})$. Then, there exists $\xi > 0$ such that

$$\|\mathbf{x} - \mathbf{x}^{\star}\| \leq \xi\, \left\|\Pi_{\mathcal{N}(\mathbf{x};\mathcal{K})}(-\mathbf{F}(\mathbf{x})) + \mathbf{F}(\mathbf{x})\right\| \quad \forall \mathbf{x} \in \mathcal{K},$$

where $\mathcal{N}(\mathbf{x};\mathcal{K}) \triangleq \left\{\mathbf{d} \in \mathbb{R}^n : \mathbf{d}^T(\mathbf{y}-\mathbf{x}) \leq 0,\ \forall \mathbf{y} \in \mathcal{K}\right\}$ denotes the normal cone.

Proof. Let $\mathbf{x} \in \mathcal{K}$ be given, and let $\mathbf{d} \triangleq \Pi_{\mathcal{N}(\mathbf{x};\mathcal{K})}(-\mathbf{F}(\mathbf{x}))$. By definition of the normal cone $\mathcal{N}(\mathbf{x};\mathcal{K})$ we have that

$$(\mathbf{y}-\mathbf{x})^T\mathbf{d} \leq 0 \quad \forall \mathbf{y} \in \mathcal{K}.$$

Since $\mathbf{x}^\star \in \mathcal{K}$, we can let $\mathbf{y} = \mathbf{x}^\star$ in the equation above, to obtain

$$(\mathbf{x}^\star - \mathbf{x})^T \left(\mathbf{d} + \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x})\right) \leq 0. \tag{3.23}$$

Since $\mathbf{x} \in \mathcal{K}$, we also have

$$\mathbf{F}(\mathbf{x}^\star)^T(\mathbf{x}^\star - \mathbf{x}) \leq 0. \tag{3.24}$$

After adding equations (3.23) and (3.24), and rearranging terms it is easy to obtain

$$\begin{aligned}(\mathbf{x}^\star - \mathbf{x})^T \left(\mathbf{F}(\mathbf{x}^\star) - \mathbf{F}(\mathbf{x})\right) &\leq (\mathbf{x} - \mathbf{x}^\star)^T \left(\mathbf{d} + \mathbf{F}(\mathbf{x})\right) \\ &\leq \|\mathbf{x} - \mathbf{x}^\star\| \left\|\mathbf{d} + \mathbf{F}(\mathbf{x})\right\|,\end{aligned} \tag{3.25}$$

where the last inequality follows from Cauchy-Schwarz inequality. By invoking the strong monotonicity of $\mathbf{F}$ we have:

$$\left(\mathbf{F}(\mathbf{x}^\star) - \mathbf{F}(\mathbf{x})\right)^T (\mathbf{x}^\star - \mathbf{x}) \geq \eta\|\mathbf{x}^\star - \mathbf{x}\|^2. \tag{3.26}$$

The desired error bound follows readily from equations (3.25) and (3.26) i.e.

$$\|\mathbf{x} - \mathbf{x}^\star\| \leq \frac{1}{\eta}\|\mathbf{d} + \mathbf{F}(\mathbf{x})\|.$$

$\square$

Proposition 3.3. Consider the VI$(\mathcal{K}, \mathbf{F})$. Let $\mathcal{K} \subset \mathbb{R}^n$ be a closed and convex set. Let the map $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be strongly monotone on $\mathbb{R}^n$ with constant $\eta > 0$ i.e. $\left(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq \eta\|\mathbf{x} - \mathbf{y}\|^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Suppose also that $\mathbf{F}$ is Lipschitz continuos with constant $L_{\mathbf{F}}$. Let $\mathbf{x}^\star$ be the unique solution of the VI$(\mathcal{K}, \mathbf{F})$. Then, there exists $\xi > 0$ such that

$$\|\mathbf{x} - \mathbf{x}^\star\| \leq \xi \left\|\mathbf{F}_{\mathcal{K}}^{nat}(\mathbf{x})\right\| \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

where $\mathbf{F}_{\mathcal{K}}^{nat}(\mathbf{x}) \triangleq \mathbf{x} - \Pi_{\mathcal{K}}(\mathbf{x} - \mathbf{F}(\mathbf{x}))$ denotes the natural map.

Proof. Let $\mathbf{x} \in \mathbb{R}^n$ be given, and let $\mathbf{v} = \mathbf{x} - \Pi_{\mathcal{K}}(\mathbf{x} - \mathbf{F}(\mathbf{x}))$. By the variational principle for the Euclidean projection, we have

$$(\mathbf{y} - \mathbf{x} + \mathbf{v})^T (\mathbf{F}(\mathbf{x}) - \mathbf{v}) \geq 0 \quad \forall \mathbf{y} \in \mathcal{K}.$$

Since $\mathbf{x}^\star \in \mathcal{K}$, we can let $\mathbf{y} = \mathbf{x}^\star$ in the equation above, to obtain

$$(\mathbf{x}^\star - \mathbf{x} + \mathbf{v})^T (\mathbf{F}(\mathbf{x}) - \mathbf{v}) \geq 0. \tag{3.27}$$

Since $\mathbf{x}^\star$ is the unique solution of the VI$(\mathcal{K}, \mathbf{F})$ and the vector $(\mathbf{x} - \mathbf{v}) \in \mathcal{K}$ we also have

$$\mathbf{F}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{v} - \mathbf{x}^\star) \geq 0. \tag{3.28}$$

After adding equations (3.27) and (3.28), and rearranging terms it is easy to obtain

$$\begin{aligned}
(\mathbf{x} - \mathbf{x}^\star)^T (\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}^\star)) &\leq \mathbf{v}^T (\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}^\star)) - \|\mathbf{v}\|^2 + (\mathbf{x} - \mathbf{x}^\star)^T \mathbf{v} \\
&\overset{(a)}{\leq} \|\mathbf{v}\| \, \|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}^\star)\| + \|\mathbf{x} - \mathbf{x}^\star\| \, \|\mathbf{v}\| \\
&\overset{(b)}{\leq} (L_{\mathbf{F}} + 1)\|\mathbf{v}\| \, \|\mathbf{x} - \mathbf{x}^\star\|
\end{aligned} \tag{3.29}$$

where (a) follows from Cauchy-Schwarz inequality, and (b) follows by the Lipchitz continuity of $\mathbf{F}$. By invoking the strong monotonicity of $\mathbf{F}$ we have:

$$(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}^\star))^T (\mathbf{x} - \mathbf{x}^\star) \geq \eta \|\mathbf{x} - \mathbf{x}^\star\|^2. \tag{3.30}$$

The desired error bound follows readily from equations (3.29) and (3.30) i.e.

$$\|\mathbf{x} - \mathbf{x}^\star\| \leq \frac{L_{\mathbf{F}} + 1}{\eta} \|\mathbf{v}\| = \frac{L_{\mathbf{F}} + 1}{\eta} \|\mathbf{F}_{\mathcal{K}}^{nat}(\mathbf{x})\|.$$

□

Note that, for a non-polyhedral finitely representable (convex) set Ξ and $\mathbf{z}$, the computation of $l_\Xi(\mathbf{z}; \mathbf{x}^\nu)$ amounts to solving a convex optimization problem; whereas in the same case if $\mathbf{z} \in \Xi$ satisfies a suitable Constraint Qualification (CQ), the computation of $r_\Xi(\mathbf{z}; \mathbf{x}^\nu)$ reduces to solving a convex *quadratic* program. Thus $r_{\Xi}(\mathbf{z};\mathbf{x}^{\nu})$ is computationally easier than $l_{\Xi}(\mathbf{z};\mathbf{x}^{\nu})$ to be obtained but, contrary to $l_{\Xi}(\mathbf{z};\mathbf{x}^{\nu})$, it can be used only to test vectors $\mathbf{z}$ belonging to Ξ. Using the error bounds in Lemma 3.2 and given a step-size rule $\{\gamma^{\nu}\}$, the termination criterion in Step 2 of the algorithm becomes then $r_{\Xi}(\mathbf{z};\mathbf{x}^{\nu}) \leq \widetilde{\alpha}\,\gamma^{\nu}$ or $l_{\Xi}(\mathbf{z};\mathbf{x}^{\nu}) \leq \widetilde{\alpha}\,\gamma^{\nu}$, for some $\widetilde{\alpha} > 0$.

3.3.2 Generalizations

So far we have restricted our attention to optimization problems in the DC form. However, it is worth mentioning that the proposed analysis and resulting algorithms (also those introduced in the forthcoming sections) can be readily extended to other sum-utility functions not necessarily in the DC form, such as $\widehat{\theta}(\mathbf{x}) \triangleq \sum_{\ell \in \mathcal{I}_f} f_{\ell}(\mathbf{x})$, where the set $I_f \triangleq \{1, \ldots, I_f\}$ may be different from the set of users $\{1, \ldots, I\}$, and each function f_{ℓ} is not necessarily expressed as the difference of two convex functions. It is not difficult to show that in such a case the candidate approximation function $\widetilde{\theta}(\mathbf{x};\mathbf{x}^{\nu})$ still has the form in (3.6), where each $\widetilde{\theta}_i(\mathbf{x};\mathbf{x}^{\nu})$ is now given by

$$\widetilde{\theta}_i(\mathbf{x};\mathbf{x}^{\nu}) \triangleq \sum_{j \in \mathcal{C}_i} f_j(\mathbf{x}_i,\, \mathbf{x}_{-i}^{\nu}) + \sum_{j \notin \mathcal{C}_i} \nabla_{\mathbf{x}_i} f_j(\mathbf{x}^{\nu})^T (\mathbf{x}_i - \mathbf{x}_i^{\nu}) + \frac{\tau_i}{2} \|\mathbf{x}_i - \mathbf{x}_i^{\nu}\|^2$$

where $\mathcal{C}_i$ is any subset of $\mathcal{S}_i \subseteq \mathcal{I}_f$, with

$$\mathcal{S}_i \triangleq \{ j \in \mathcal{I}_f : f_j(\bullet, \mathbf{x}_{-i}) \text{ is convex on } \mathcal{X}_i, \forall \mathbf{x}_{-i} \in \mathcal{X}_{-i} \}\,.$$

In $\widetilde{\theta}_i(\mathbf{x};\mathbf{x}^{\nu})$ each user linearizes only the functions outside $\mathcal{C}_i$ while preserving the convex part of the sum-utility. The choice of $\mathcal{C}_i \subseteq \mathcal{S}_i$ is a degree of freedom useful to explore the tradeoff between signaling and convergence speed.

3.4 Distributed Implementation

In general, the implementation of Algorithms 3.1 and 3.2 requires a coordination among the users; the amount of network signaling depends on the specific application under consideration. To alleviate the communication overhead of a centralized implementation, it is desirable to obtain a decentralized version of these schemes. Interestingly, the separability structure of the approximation function $\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu)$ resulting from the proposed convexification method (cf. Section 3.3) as well as that of the coupling constraints (cf. A5) lends itself to a parallel decomposition of the subproblems (3.7) across the users in the primal or dual domain. The proposed distributed implementations of Step 2 of Algorithms 3.1 (and Algorithm 3.2) are described in the next two subsections.

3.4.1 Distributed Dual-Decomposition based Algorithms

The subproblems (3.7) can be solved in a distributed way if the side constraints $\mathbf{h}(\mathbf{x}) \leq \mathbf{0}$ are dualized (under zero-duality gap). The dual problem associated with each (3.7) is: given $\mathbf{x}^\nu \in \Xi$,

$$\underset{\boldsymbol{\lambda}\geq\mathbf{0}}{\text{maximize}} \left\{ d(\boldsymbol{\lambda};\mathbf{x}^\nu) \triangleq \min_{\mathbf{x}\in\mathcal{X}} \left\{ \widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu) + \boldsymbol{\lambda}^T\mathbf{h}(\mathbf{x}) \right\} \right\}. \tag{3.31}$$

Note that, under A1-A5, the inner minimization in (3.31) has a unique solution, which will be denoted by $\widehat{\mathbf{x}}(\boldsymbol{\lambda};\mathbf{x}^\nu) \triangleq (\widehat{\mathbf{x}}_i(\boldsymbol{\lambda};\mathbf{x}^\nu))_{i=1}^I$, that is

$$\widehat{\mathbf{x}}_i(\boldsymbol{\lambda};\mathbf{x}^\nu) \triangleq \underset{\mathbf{x}_i\in\mathcal{X}_i}{\operatorname{argmin}} \left\{ \widetilde{\theta}_i(\mathbf{x}_i;\mathbf{x}^\nu) + \boldsymbol{\lambda}^T\mathbf{h}_i(\mathbf{x}_i) \right\}. \tag{3.32}$$

Before proceeding, let us introduce the following assumptions.

A6) The side constraint vector function $\mathbf{h}(\bullet)$ is Lipschitz continuous on $\mathcal{X}$, with constant $L_{\mathbf{h}}$.

A7) For each subproblem (3.7), there is zero-duality gap, and the dual problem (3.31) has a non-empty solution set.

We emphasize that the above conditions are generally satisfied by many practical problems of interest. For example, A7 holds if some CQ is satisfied, e.g., (generalized) Slater's CQ, or the feasible set Ξ is a polyhedron.

The next lemma summarizes some desirable properties of the dual function $d(\boldsymbol{\lambda};\mathbf{x}^\nu)$, which are instrumental to prove convergence of dual schemes.

Lemma 3.3. Under A1-A5 we have the following:

(a) $d(\boldsymbol{\lambda};\mathbf{x}^\nu)$ is differentiable on $\mathbb{R}_+^{n_c}$, with gradient

$$\nabla_{\boldsymbol{\lambda}} d(\boldsymbol{\lambda};\mathbf{x}^\nu) = \mathbf{h}\left(\widehat{\mathbf{x}}(\boldsymbol{\lambda};\mathbf{x}^\nu)\right) = \sum_i \mathbf{h}_i\left(\widehat{\mathbf{x}}_i(\boldsymbol{\lambda};\mathbf{x}^\nu)\right). \tag{3.33}$$

(b) If, in addition, A6 holds, then $\nabla_{\boldsymbol{\lambda}} d(\boldsymbol{\lambda};\mathbf{x}^\nu)$ is Lipschitz continuous on $\mathbb{R}_+^{n_c}$ with constant $L_{\nabla d} \triangleq L_{\mathbf{h}}^2\sqrt{n_c}/c_{\boldsymbol{\tau}}$.

Proof. (*a*) It is a consequence of Danskin's theorem [12, Prop. A.43].
(b) The statement follows from

$$\begin{aligned}\|\nabla d\left(\boldsymbol{\lambda};\widehat{\mathbf{x}}\left(\boldsymbol{\lambda};\mathbf{x}^\nu\right)\right) - \nabla d(\boldsymbol{\lambda};\widehat{\mathbf{x}}(\boldsymbol{\lambda}';\mathbf{x}^\nu))\| &\le L_{\mathbf{h}}\|\widehat{\mathbf{x}}\left(\boldsymbol{\lambda};\mathbf{x}^\nu\right) - \widehat{\mathbf{x}}(\boldsymbol{\lambda}';\mathbf{x}^\nu)\| \\ &\le L_{\hat{x}}\|\boldsymbol{\lambda}-\boldsymbol{\lambda}'\|,\end{aligned}$$

where in the last inequality we used the uniformly Lipschitz property of $\widehat{\mathbf{x}}(\bullet;\mathbf{x}^\nu)$, as proved in the lemma below. $\square$

Lemma 3.4. Under A1-A6, $\widehat{\mathbf{x}}(\bullet;\mathbf{x}^\nu)$ is uniformly Lipschitz continuous on $\mathbb{R}_+^{n_c}$, with constant $L_{\hat{x}} \triangleq L_{\mathbf{h}}\sqrt{n_c}/c_{\boldsymbol{\tau}}$.

Proof. To simplify the notation, let us write $\mathbf{x}_{\boldsymbol{\lambda}} \triangleq \widehat{\mathbf{x}}\left(\boldsymbol{\lambda};\mathbf{x}^\nu\right)$ and $\mathbf{x}_{\boldsymbol{\lambda}'} \triangleq \widehat{\mathbf{x}}\left(\boldsymbol{\lambda}';\mathbf{x}^\nu\right)$. By the minimum principle, we have that

$$\begin{aligned}\left(\mathbf{x}_{\boldsymbol{\lambda}'} - \mathbf{x}_{\boldsymbol{\lambda}}\right)^T \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}},\boldsymbol{\lambda}) \ge 0 \\ \left(\mathbf{x}_{\boldsymbol{\lambda}} - \mathbf{x}_{\boldsymbol{\lambda}'}\right)^T \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}'},\boldsymbol{\lambda}') \ge 0,\end{aligned}$$

where $\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) \triangleq \widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu) + \sum_{j=1}^{n_c}\lambda_j h_j(\mathbf{x})$ denotes the Lagrangian function. Adding these two inequalities, we obtain

$$\left(\mathbf{x}_{\boldsymbol{\lambda}} - \mathbf{x}_{\boldsymbol{\lambda}'}\right)^T\left[\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}'},\boldsymbol{\lambda}') - \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}},\boldsymbol{\lambda}') + \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}},\boldsymbol{\lambda}') - \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}},\boldsymbol{\lambda})\right] \ge 0,$$

which leads to

$$\begin{aligned}c_{\boldsymbol{\tau}}\|\mathbf{x}_{\boldsymbol{\lambda}'} - \mathbf{x}_{\boldsymbol{\lambda}}\|^2 &\le \left(\mathbf{x}_{\boldsymbol{\lambda}'} - \mathbf{x}_{\boldsymbol{\lambda}}\right)^T\left[\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}'},\boldsymbol{\lambda}') - \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}},\boldsymbol{\lambda}')\right] \\ &\le \left(\mathbf{x}_{\boldsymbol{\lambda}} - \mathbf{x}_{\boldsymbol{\lambda}'}\right)^T\left[\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}},\boldsymbol{\lambda}') - \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}},\boldsymbol{\lambda})\right],\end{aligned}$$

where the first inequality follows by the uniform strong convexity of $\mathcal{L}(\bullet,\boldsymbol{\lambda}')$ (with constant $c_{\boldsymbol{\tau}}$). Then, it is not difficult to see that

$$
\begin{aligned}
c_{\boldsymbol{\tau}} \left\| \mathbf{x}_{\boldsymbol{\lambda}'} - \mathbf{x}_{\boldsymbol{\lambda}} \right\| & \overset{(a)}{\leq} \left\| \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}}, \boldsymbol{\lambda}') - \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}_{\boldsymbol{\lambda}}, \boldsymbol{\lambda}) \right\| \\
& \overset{(b)}{\leq} \sum_{j=1}^{n_c} \left| \lambda_j' - \lambda_j \right| \left\| \nabla_{\mathbf{x}} h_j(\mathbf{x}_{\boldsymbol{\lambda}}) \right\| \\
& \overset{(c)}{\leq} \sum_{j=1}^{n_c} \left| \lambda_j' - \lambda_j \right| L_{\mathbf{h}} = L_{\mathbf{h}} \left\| \boldsymbol{\lambda}' - \boldsymbol{\lambda} \right\|_1 \\
& \overset{(d)}{\leq} L_{\mathbf{h}} \sqrt{n_c} \left\| \boldsymbol{\lambda}' - \boldsymbol{\lambda} \right\|
\end{aligned}
$$

where: (a) follows by Cauchy-Schwartz inequality, (b) is a consequence of the triangle inequality, (c) is a direct consequence of the Lipschitz continuity of the $\mathbf{h}$, and (d) follows by the equivalence of the norms. $\square$

The dual-problem can be solved, e.g., using well-known gradient algorithms [11]; an instance is given in Algorithm 3.3, whose convergence is stated in Theorem 3.4. In (3.34) $[\bullet]^+$ denotes the Euclidean projection onto $\mathbb{R}_+$, i.e., $[x]^+ \triangleq \max(0, x)$.

Algorithm 3.3: Dual-based Distributed Implementation of Step 2 of Algorithm 3.1

Data: $\boldsymbol{\lambda}^0 \geq \mathbf{0}$, $\mathbf{x}^\nu$, $\{\alpha^t\} > 0$; set $t = 0$.

(`S.2a`): If $\boldsymbol{\lambda}^t$ satisfies a suitable termination criterion: STOP.

(`S.2b`): The users solve in parallel (3.32): for all $i = 1, \ldots, I$, compute $\widehat{\mathbf{x}}_i(\boldsymbol{\lambda}^t; \mathbf{x}^\nu)$.

(`S.2c`): Update $\boldsymbol{\lambda}$ according to

$$
\boldsymbol{\lambda}^{t+1} \triangleq \left[\boldsymbol{\lambda}^t + \alpha^t \sum_{i=1}^{I} \mathbf{h}_i \left(\widehat{\mathbf{x}}_i(\boldsymbol{\lambda}^t; \mathbf{x}^\nu) \right) \right]^+ . \tag{3.34}
$$

(`S.2d`): $t \leftarrow t + 1$ and go back to (`S.2a`).

Theorem 3.4. Given the DC program (3.1) under A1-A5, suppose that one of the two following conditions are satisfied:

(a) A6 holds and $\{\alpha^t\}$ is chosen such that $0 < \alpha^t = \alpha^{\max} < 2/L_{\nabla d}$, for all $t \geq 0$;

(b) $\nabla_{\boldsymbol{\lambda}} d(\bullet; \mathbf{x}^\nu)$ is uniformly bounded on $\mathbb{R}_+^{n_c}$, and $\{\alpha^t\}$ is chosen such that $\alpha^t > 0$, $\alpha^t \to 0$, $\sum_t \alpha^t = \infty$, and $\sum_t (\alpha^t)^2 < \infty$.

Then, under A7, the sequence $\{\boldsymbol{\lambda}^t\}$ generated by Algorithm 3.3 converges to a solution of (3.31), and the sequence $\{\widehat{\mathbf{x}}(\boldsymbol{\lambda}^t; \mathbf{x}^\nu)\}$ converges to the unique solution of (3.7).

Proof. This result is a direct consequence of Lemma 3.3 and standard convergence results of gradient projection algorithms; refer to [105, Thm. 3.2] and [11, Prop. 8.2.6] for the conclusions stated under assumptions (a) and (b), respectively. □

Remark 3.2 (On the distributed implementation)**.** The distributed implementation of Algorithms 3.1 and 3.2 based on Algorithm 3.3 leads to a double-loop scheme with communication between the two loops: given the current value of the multipliers $\boldsymbol{\lambda}^t$, the users can solve in a distributed way their subproblems (3.32); once the new value $\widehat{\mathbf{x}}(\boldsymbol{\lambda}^t; \mathbf{x}^\nu)$ is available, the multipliers are updated according to (3.34). Note that when $n_c = 1$ (i.e., only one shared constraint), the update in (3.34) can be replaced by a bisection search, which generally converges quite fast. When $n_c > 1$, the potential slow convergence of gradient updates (3.34) can be alleviated using accelerated gradient-based update; see, e.g., [77]. Note also that the size of the dual problem (the dimension of $\boldsymbol{\lambda}$) is equal to n_c (the number of shared constraints), which makes Algorithm 3.3 scalable in the number of users.

As far as the communication overhead needed to implement the proposed scheme is concerned, the signaling among the users is in the form of message passing and, of course, is problem dependent; see Section 3.5 for specific examples. When the networks has a cluster-head, the update of the multipliers can be performed at the cluster, and then broadcast to the users. In fully decentralized networks, the update of $\boldsymbol{\lambda}$ can be done by the users themselves, by running consensus based algorithms to locally estimate $\sum_i \mathbf{h}_i\left(\widehat{\mathbf{x}}_i(\boldsymbol{\lambda}^t; \mathbf{x}^\nu)\right)$. This in general requires a limited signaling exchange among neighboring nodes.

3.4.2 Distributed Implementation via Primal-Decomposition

Algorithm 3.3 is based on the relaxation of the side constraints into the Lagrangian, resulting in general in a violation of these constraints during the intermediate iterates. In some applications, this may prevent the on-line implementation of the algorithm. In this section, we propose a distributed scheme which does not suffer from this issue; we cope with side constraints using a primal decomposition technique.

Introducing the slack variables $\mathbf{t} \triangleq (\mathbf{t}_i)_{i=1}^I$, with each $\mathbf{t}_i \in \mathbb{R}^{n_c}$, (3.7) can be rewritten as

$$\begin{array}{lll} \underset{(\mathbf{x}_i,\mathbf{t}_i)_{i=1}^I}{\text{minimize}} & \displaystyle\sum_{i=1}^{I} \widetilde{\theta}_i(\mathbf{x}_i;\mathbf{x}^\nu), & \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i, & \forall i = 1,\ldots,I, \\ & \mathbf{h}_i(\mathbf{x}_i) \le \mathbf{t}_i, & \forall i = 1,\ldots,I, \\ & \displaystyle\sum_{i=1}^{I} \mathbf{t}_i \le \mathbf{0}, & \end{array} \tag{3.35}$$

When $\mathbf{t} = (\boldsymbol{t}_i)_{i=1}^I$ is fixed, (3.35) can be decoupled across the users: for each $i = 1,\ldots,I$, solve

$$\begin{array}{ll} \underset{\mathbf{x}_i}{\text{minimize}} & \widetilde{\theta}_i(\mathbf{x}_i;\mathbf{x}^\nu), \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i, \\ & \mathbf{h}_i(\mathbf{x}_i) \overset{\boldsymbol{\mu}_i(\mathbf{t}_i;\mathbf{x}^\nu)}{\le} \mathbf{t}_i, \end{array} \tag{3.36}$$

where $\boldsymbol{\mu}_i(\mathbf{t}_i;\mathbf{x}^\nu)$ is the optimal Lagrange multiplier associated with the inequality constraint $\mathbf{h}_i(\mathbf{x}_i) \le \mathbf{t}_i$. Note that the existence of $\boldsymbol{\mu}_i(\mathbf{t}_i;\mathbf{x}^\nu)$ is guaranteed if (3.36) has zero duality gap (e.g., when some CQ hold) [11, Prop. 6.5.8], but $\boldsymbol{\mu}_i(\mathbf{t}_i;\mathbf{x}^\nu)$ may not be unique. Let us denote by $\mathbf{x}_i^\star(\mathbf{t}_i;\mathbf{x}^\nu)$ the unique solution of (3.36) given $\mathbf{t} = (\boldsymbol{t}_i)_{i=1}^I$. The optimal slack $\mathbf{t}^\star \triangleq (\mathbf{t}_i^\star)_{i=1}^I$ of the shared constraints can be found solving the so-called *master* (convex) problem (see, e.g., [81]):

$$\begin{array}{ll} \underset{\boldsymbol{t}}{\text{minimize}} & P(\boldsymbol{t};\mathbf{x}^\nu) \triangleq \displaystyle\sum_{i=1}^{I} \widetilde{\theta}_i(\mathbf{x}_i^\star(\mathbf{t}_i;\mathbf{x}^\nu);\mathbf{x}^\nu) \\ \text{subject to} & \displaystyle\sum_{i=1}^{I} \mathbf{t}_i \le \mathbf{0}. \end{array} \tag{3.37}$$

Due to the non-uniqueness of $\boldsymbol{\mu}_i(\mathbf{t}_i;\mathbf{x}^\nu)$, the objective function in (3.37) is nondifferentiable; problem (3.37) can be solved by subgradient methods. A subgradient of $P(\boldsymbol{t};\mathbf{x}^\nu)$ at $\mathbf{t}$ is

$$\partial_{\mathbf{t}_i} P(\mathbf{t};\mathbf{x}^\nu) = -\boldsymbol{\mu}_i(\mathbf{t}_i;\mathbf{x}^\nu), \quad i = 1,\ldots,I.$$

We refer to [11, Prop. 8.2.6] for standard convergence results of subgradient projection algorithms.

3.5 Applications and Numerical Results

The DC formulation (3.1) and the consequent proposed algorithms are general enough to encompass many problems of practical interest in different fields. Here we focus on two specific applications that can be casted in the DC-form (3.1), namely: i) a new secrecy rate maximization game; and ii) the sum-rate maximization problem over MIMO CR networks.

3.5.1 Case Study 1: A New Secrecy Rate Game

Physical layer security has been considered as a promising technique to prevent illegitimate receivers from eavesdropping on the confidential message transmitted between intended network nodes; see, e.g., [47] and references therein. Recently, cooperative transmissions using trusted relays or friendly jammers to improve physical layer security has attracted increasing attention [24, 39, 59, 115, 104, 36, 119, 60]. There are mainly three relaying/jamming protocols for the cooperative secure transmission: amplify-and-forward, decode-and-forward, and cooperative (friendly) jamming. Of particular interest to this work is the Cooperative Jamming (CJ) paradigm (see, e.g., [24, 39, 59]): friendly jammers create judicious interference by transmitting noise (or codewords) so as to impair the eavesdropper's ability to decode the confidential information, and thus, increase secure communication rates between legitimate parties. The interference from the jammers however might also reduce the useful rate of the legitimate links; therefore the maximization of the users' secrecy rate calls for a *joint* optimization of the power allocations of the sources *and* the jammers.

Here we address such a joint optimization problem. We consider a network model composed of *multiple* transmitter-receiver pairs, *multiple* friendly jammers, and a single eavesdropper. Note that previous works studied simpler system models, composed of either *one* source-destination link (and possibly multiple jammers) or multiple sources but *one* jammer. We formulate the system design as a game wherein the players—the legitimate users—maximize their own secrecy rate by choosing *jointly* their transmit power and (the optimal fraction of the) power of the friendly jammers. The resulting secrecy game faces two main challenges, namely: i) the players' objective functions are nonconcave and nondifferentiable; and ii) there are side (thus coupling) constraints. All this makes the analysis of the proposed game a difficult task; for instance, a Nash Equilibrium may not even exist. Capitalizing on recent results on nonconvex games with side constraints [84, 85], we introduce a novel relaxed equilibrium concept for the nonconvex nondifferentiable game, named (restricted) B-Quasi Generalized Nash Equilibrium (B-QGNE). Roughly speaking a B-QGNE is a solution of the first order aggregated stationarity conditions (based on directional derivatives) of the players' optimization problems. Aiming to devise distributed algorithms computing a (B-)QGNE, we establish a connection between (a subclass of) such equilibria and the stationary solutions of a suitably defined differentiable DC program with side constraints, for which we can successfully use the DC framework developed in this this chapter. To the best of our knowledge, this is the first attempt toward a rigorous characterization of the nondifferentiability issue in secrecy rate multiuser resource allocation problems. We conclude our analysis with some numerical experiments, which show that the proposed *distributed* algorithms yield secrecy sum-rates that are better than those achievable by *centralized* schemes (attempting to compute stationary solutions of the DC program), and comparable to those achievable by computationally expensive techniques attempting to obtain globally optimal solutions.

3.5.1.1 System Model and Problem Formulation

Consider a wireless communication system composed of Q transmitter-receiver pairs—the legitimate users—J friendly jammers, and a single eavesdropper; see Figure 3.1. We assume OFDMA transmissions for the legitimate users over flat-fading and quasi-static (constant within the transmission frame) channels. We denote by $H_{qq}^{\mathtt{SD}}$ the channel gain of the legitimate source-destination pair q, by $H_{jq}^{\mathtt{JD}}$ the channel gain between the transmitter of jammer j and the receiver of user q, by $H_{je}^{\mathtt{JE}}$ the channel gain between the transmitter of jammer j and the receiver of the eavesdropper, and by $H_{qe}^{\mathtt{SE}}$ the channel between the source q and the eavesdropper. We assume CSI of the eavesdropper's (cross)-channels; this is a common assumption in PHY security literature; see, e.g., [47, 24, 39, 59, 115, 104, 119, 36, 60]. CSI on the eavesdropper's channel can be obtained when the eavesdropper is active in the network and its transmissions can be monitored. For instance, this happens in networks combining multicast and unicast transmissions wherein the users work as legitimate receiver for some signals and eavesdroppers for others. Another scenario is a cellular environment where the eavesdroppers are also users of the network; in a time-division duplex system, the base station can estimate the users' channels based on the channel reciprocity.

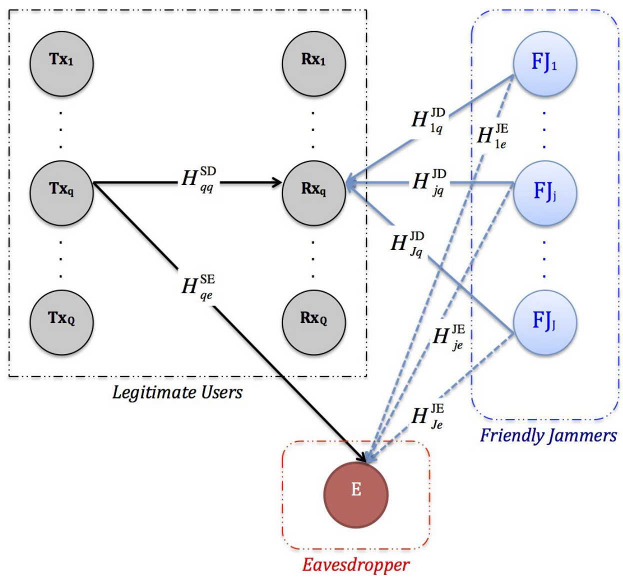

Figure 3.1: System model composed of the legitimate users, friendly jammers and a single eavesdropper. The arrows illustrate the channel gains for the q-th user and the eavesdropper.

In the setting above, we adopt the CJ paradigm: the friendly jammers cooperate with the users by introducing a proper interference profile "masking" the eavesdropper. The (uniform) power allocation of source q is denoted by p_q; $p^{\mathtt{J}}_{jq}$ is the fraction of power of friendly jammer j requested by user q (allocated by jammer j over the channel used by user q); the power profile allocated by all the jammers over the channel of user q is $\mathbf{p}^{\mathtt{J}}_q \triangleq (p^{\mathtt{J}}_{jq})_{j=1}^{J}$. Each user q has power budget $p_q \leq P_q$, and likewise each jammer j, that is $\sum_{q=1}^{Q} p^{\mathtt{J}}_{jq} \leq P^{\mathtt{J}}_j$, for all $j = 1, \ldots, J$. Under basic information theoretical assumptions, the maximum achievable rate on link q is

$$r_{qq}\left(p_q, \mathbf{p}^{\mathtt{J}}_q\right) \triangleq \log\left(1 + \frac{H^{\mathtt{SD}}_{qq}\, p_q}{\sigma^2 + \sum_{j=1}^{J} H^{\mathtt{JD}}_{jq}\, p^{\mathtt{J}}_{jq}}\right). \tag{3.38}$$

Similarly, the rate on the channel between source q and the eavesdropper is

$$r_{qe}\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right) \triangleq \log\left(1 + \frac{H_{qe}^{\mathtt{SE}}\, p_q}{\sigma^2 + \sum_{j=1}^{J} H_{je}^{\mathtt{JE}}\, p_{jq}^{\mathtt{J}}}\right). \tag{3.39}$$

The secrecy rate of user q is then (see, e.g., [47]):

$$r_q^s\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right) \triangleq \left[r_{qq}(p_q, \mathbf{p}_q^{\mathtt{J}}) \,-\, r_{qe}(p_q, \mathbf{p}_q^{\mathtt{J}})\right]^+. \tag{3.40}$$

Problem Formulation. We formulate the system design as a game where the legitimate users are the players who cooperate with the jammers to maximize their own secrecy rate. More formally, anticipating $(\mathbf{p}_r^{\mathtt{J}})_{r\neq q}$, each user q seeks together with the jammers the tuple $\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right)$ solving the following optimization problem:

$$\mathcal{G}: \quad \begin{array}{ll} \underset{\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right)\geq \mathbf{0}}{\text{maximize}} & r_q^s\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right) \\ \text{subject to:} & \\ & \left.\begin{array}{l} p_q \leq P_q, \\ \displaystyle\sum_{r=1}^{Q} p_{jr}^{\mathtt{J}} \leq P_j^{\mathtt{J}}, \ \forall\, j = 1, \ldots, J. \end{array}\right\} \triangleq \mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}}) \end{array} \tag{3.41}$$

Note that the feasible set $\mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}})$ of (3.41) depends on the jammers' power profile $\mathbf{p}_{-q}^{\mathtt{J}} \triangleq (\mathbf{p}_r^{\mathtt{J}})_{r\neq q}$ allocated over the other users' channels. When needed, we will denote each tuple $(p_q, \mathbf{p}_q^{\mathtt{J}})$ by $\mathbf{x}_q \triangleq (p_q, \mathbf{p}_q^{\mathtt{J}})$. The game whose q-th optimization problem is given by (3.41) will be termed as secrecy game $\mathcal{G}$.

The secrecy game $\mathcal{G}$ is an instance of the so-called Generalized Nash Equilibrium problem (GNEP) with shared constraints; see, e.g., [27]. A solution of $\mathcal{G}$ is the Generalized Nash Equilibrium (GNE), defined as follows.

Definition 3.1 (GNE)**.** A strategy profile $(p_q^\star, \mathbf{p}_q^{\mathtt{J}\star})_{q=1}^{Q}$ is a GNE of the GNEP $\mathcal{G}$ if the following holds for all $q = 1, \ldots, Q$: $(p_q^\star, \mathbf{p}_q^{\mathtt{J}\star}) \in \mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}\star})$ and

$$r_q^s(p_q^\star, \mathbf{p}_q^{\mathtt{J}\star}) \geq r_q^s(p_q, \mathbf{p}_q^{\mathtt{J}}), \quad \forall (p_q, \mathbf{p}_q^{\mathtt{J}}) \in \mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}\star}). \tag{3.42}$$

The (distributed) computation of a GNE of $\mathcal{G}$ is a challenging if not an impossible task, due to the following issues:

1. The nondifferentiability of the player's objective functions.

2. The nonconcavity of the players' objective functions.

3. The presence of coupling constraints.

Toward the practical resolution of $\mathcal{G}$ coping with the aforementioned issues, we follow the 3 logical steps outlined next.

- *A smooth-game formulation*: We start dealing with the nondifferentiability issue by introducing a smooth restricted (still nonconcave) version of the original game $\mathcal{G}$, termed game $\mathcal{G}^{\mathtt{sm}}$, and establishing the connection with $\mathcal{G}$ in terms of GNE;

- *Relaxed equilibrium concepts*: To deal with the nonconcavity of the players' objective functions we introduce relaxed equilibrium concepts for both games $\mathcal{G}$ and $\mathcal{G}^{\mathtt{sm}}$, and establish their connections. The comparison shows that the smooth game $\mathcal{G}^{\mathrm{sm}}$ preserves (relaxed) solutions of $\mathcal{G}$ of practical interest while just ignoring those yielding zero secrecy rates for the players.

- *Algorithmic design*: We then focus on the computation of the relaxed equilibria of $\mathcal{G}^{\mathtt{sm}}$, casting the problem into a multiuser DC program in the form (3.1), which can be distributively solved using the machinery introduced earlier in this chapter.

3.5.1.2 A Smooth-Game Formulation

We introduce a restricted smooth-game wherein the max operator in the objective function of each player's optimization problem (3.41) is relaxed via linear constraints. This formulation is a direct consequence of the following fact: $r_{qq}\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right) \geq r_{qe}\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right)$ if and only if either $p_q = 0$ or

$$\frac{H_{qq}^{\mathtt{SD}}}{\sigma^2 + \sum_{j=1}^{J} H_{jq}^{\mathtt{JD}}\, p_{jq}^{\mathtt{J}}} \geq \frac{H_{qe}^{\mathtt{SE}}}{\sigma^2 + \sum_{j=1}^{J} H_{je}^{\mathtt{JE}}\, p_{jq}^{\mathtt{J}}}.$$

Clearly, the latter inequality is equivalent to:

$$\sum_{j=1}^{J} \left(H_{qq}^{\mathtt{SD}} H_{je}^{\mathtt{JE}} - H_{qe}^{\mathtt{SE}} H_{jq}^{\mathtt{JD}}\right) p_{jq}^{\mathtt{J}} + \left(H_{qq}^{\mathtt{SD}} - H_{qe}^{\mathtt{SE}}\right) \sigma^2 \geq 0. \tag{3.43}$$

Note that if (3.43) holds with equality, then $r_q^s\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right) = 0$ for any $p_q \geq 0$. These observations lead to the restricted smooth game, which we call $\mathcal{G}^{\mathtt{sm}}$, where each player q anticipating $\mathbf{p}_{-q}^{\mathtt{J}}$ solves the following differentiable, albeit nonconcave, maximization problem:

$$\mathcal{G}^{\mathtt{sm}}: \quad \begin{array}{ll} \underset{\left(p_q, \mathbf{p}_q^{\mathtt{J}}\right) \geq \mathbf{0}}{\text{maximize}} & \widetilde{r}_q^s(p_q, \mathbf{p}_q^{\mathtt{J}}) \triangleq r_{qq}(p_q, \mathbf{p}_q^{\mathtt{J}}) \, - \, r_{qe}(p_q, \mathbf{p}_q^{\mathtt{J}}) \\ \text{subject to:} & \\ & \left.\begin{array}{l} p_q \leq P_q, \\ \text{constraint (3.43)}, \\ \displaystyle\sum_{r=1}^{Q} p_{jr}^{\mathtt{J}} \leq P_j^{\mathtt{J}}, \forall\, j = 1, \ldots, J, \end{array}\right\} \triangleq \mathcal{P}_q^{\mathtt{sm}}(\mathbf{p}_{-q}^{\mathtt{J}}) \end{array} \tag{3.44}$$

where we denoted by $\mathcal{P}_q^{\mathrm{sm}}(\mathbf{p}_{-q}^{\mathtt{J}})$ the feasible set of the optimization problem. For notational convenience, we also introduce the joint strategy set $\mathcal{P}$ defined as:

$$\mathcal{P} \triangleq \left\{ \begin{array}{lcl} (p_q, \mathbf{p}_q^{\mathtt{J}}))_{q=1}^{Q} \geq \mathbf{0} & : & p_q \leq P_q \text{ and (3.43) holds, } \forall q = 1, \ldots, Q, \\ & & \sum_{r=1}^{Q} p_{jr}^{\mathtt{J}} \leq P_j^{\mathtt{J}}, \quad \forall\, j = 1, \ldots, J \end{array} \right\}.$$

It turns out from the above discussion that, solution-wise, the main difference between the smooth game $\mathcal{G}^{\mathtt{sm}}$ and the original one $\mathcal{G}$ is that $\mathcal{G}^{\mathtt{sm}}$ ignores the feasible players' strategy profiles of $\mathcal{G}$ violating (3.43) (and thus outside $\mathcal{P}$). But such tuples yield zero secrecy rate of the players and thus are of little significance, since the players' goals are to attempt the maximization of their secrecy rates. We can then focus on strategy profiles in the set $\mathcal{P}$, without any practical loss of optimality. The next proposition makes formal the aforementioned connection between $\mathcal{G}$ and $\mathcal{G}^{\mathtt{sm}}$.

Proposition 3.4. Given $\mathcal{G}$ and $\mathcal{G}^{\mathtt{sm}}$, the following hold.

(a) A GNE of $\mathcal{G}$ always exists;

(b) A GNE of $\mathcal{G}^{\mathtt{sm}}$ exists provided that $\mathcal{P} \neq \emptyset$;

(c) $[\mathcal{G} \to \mathcal{G}^{\mathtt{sm}}]$: If $\mathbf{x}^\star$ is a GNE of $\mathcal{G}$ satisfying the constraints (3.43) for all $q = 1, \ldots, Q$, then $\mathbf{x}^\star$ is a GNE of $\mathcal{G}^{\mathtt{sm}}$.

(d) $[\mathcal{G}^{\mathtt{sm}} \to \mathcal{G}]$: If $\mathbf{x}^\star$ is a GNE of $\mathcal{G}^{\mathtt{sm}}$, then $\mathbf{x}^\star$ is a GNE of $\mathcal{G}$.

Proof. (a) It is not difficult to check that $\mathcal{G}$ is an exact potential game with potential function $\Phi(\mathbf{p}, \mathbf{p}^{\mathtt{J}}) \triangleq \sum_q r_q^s(p_q, \mathbf{p}_q^{\mathtt{J}})$. It turns out that any optimal solution of the associated multiplayer maximization problem:

$$\begin{array}{ll} \underset{(\mathbf{p},\mathbf{p}^{\mathtt{J}})\geq \mathbf{0}}{\text{maximize}} & \Phi(\mathbf{p}, \mathbf{p}^{\mathtt{J}}) \\ \text{subject to:} & \begin{array}{ll} p_q \leq P_q, & \forall q = 1, \ldots, Q \\ \displaystyle\sum_{r=1}^{Q} p_{jr}^{\mathtt{J}} \leq P_j^{\mathtt{J}}, & \forall j = 1, \ldots, J, \end{array} \end{array} \tag{3.45}$$

is a GNE of $\mathcal{G}$. Since (3.45) has a solution, there must exist a GNE for $\mathcal{G}$.

(b) The proof is based on similar arguments as those in (a).

(c) Suppose that $\mathbf{x}^{\star} \triangleq \left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right)_{q=1}^{Q}$ is a GNE of $\mathcal{G}$ satisfying (3.43) for all $q = 1, \ldots, Q$. Then, for every q, we have: $\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right) \in \mathcal{P}_q^{\mathtt{sm}}(\mathbf{p}_{-q}^{\mathtt{J}\star})$ and

$$\widetilde{r}_q^s\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right) = r_q^s\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right) \geq r_q^s\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right) = \widetilde{r}_q^s\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right),$$

for all $\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right) \in \mathcal{P}_q^{\mathtt{sm}}(\mathbf{p}_{-q}^{\mathtt{J}\star})$. Therefore, $\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right)$ is a solution of (3.44), with $\mathbf{p}_{-q}^{\mathtt{J}} = \mathbf{p}_{-q}^{\mathtt{J}\star}$.

(d) Suppose that $\mathbf{x}^{\star} \triangleq \left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right)_{q=1}^{Q}$ is a GNE of $\mathcal{G}^{\mathtt{sm}}$. Then, for each q, $\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right) \in \mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}\star})$ and

$$r_q^s\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right) = \widetilde{r}_q^s\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right) \geq \widetilde{r}_q^s\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right) = r_q^s\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right),$$

for all $\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right) \in \mathcal{P}_q^{\mathtt{sm}}(\mathbf{p}_{-q}^{\mathtt{J}\star})$. Since $r_q^s\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right) = 0$ for all $\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right) \in \mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}\star}) \setminus \mathcal{P}_q^{\mathtt{sm}}(\mathbf{p}_{-q}^{\mathtt{J}\star})$, we have

$$r_q^s\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right) \geq r_q^s\left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right), \quad \forall \left(\mathbf{p}_q, \mathbf{p}_q^{\mathtt{J}}\right) \in \mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}\star}).$$

Therefore, $\left(\mathbf{p}_q^{\star}, \mathbf{p}_q^{\mathtt{J}\star}\right)$ is a solution of (3.41), with $\mathbf{p}_{-q}^{\mathtt{J}} = \mathbf{p}_{-q}^{\mathtt{J}\star}$. □

The first important result stated in Proposition 3.4 is that the two games have a solution ($\mathcal{G}^{\mathtt{sm}}$ under $\mathcal{P} \neq \emptyset$). Note that a sufficient condition guaranteeing $\mathcal{P} \neq \emptyset$ is $H_{qq}^{\mathtt{SD}} \geq H_{qe}^{\mathtt{SE}}$ for all q [cf. (3.43)], implying that the channel gains of the legitimate users cannot be worse than those between the sources and the eavesdropper. For instance, this happens if the legitimate receivers are much closer to their intended transmitters than the eavesdropper's receiver. Conversely, if (3.43) is violated for some q (implying $\mathcal{P} = \emptyset$), there exists no feasible user/jammer power allocation yielding a positive secrecy rate for user q. Note that this is in agreement with current results in the literature; see, e.g., [36, 104].

Proposition 3.4 also establishes the connection between the GNE of $\mathcal{G}$ and $\mathcal{G}^{\mathtt{sm}}$. In particular, statement (d) justifies the focus on the smooth (and thus more affordable) game $\mathcal{G}^{\mathtt{sm}}$ rather than the nonsmooth $\mathcal{G}$ without any practical loss of generality. The computation of a GNE of $\mathcal{G}^{\mathtt{sm}}$ however remains a difficult task, because of the nonconcavity of the players' objective functions $\widetilde{r}_q^s$. To obtain practical solution schemes, we introduce next a relaxed equilibrium concept, whose computation can be done using the framework proposed in the first part of this chapter.

3.5.1.3 Relaxed Equilibrium Concepts

Based on the concept of B(ouligand)-derivative [28] we define next a relaxed notion of equilibrium for the nonsmooth nonconvex game $\mathcal{G}$.

Definition 3.2 (B-QGNE)**.** A strategy profile $\mathbf{x}^\star \triangleq (\mathbf{x}_q^\star \triangleq (p_q^\star, \mathbf{p}_q^{\mathtt{J}\star}))_{q=1}^Q$ is a B-Quasi GNE (B-QGNE) of $\mathcal{G}$ if the following holds for all $q = 1, \dots, Q$: $(p_q^\star, \mathbf{p}_q^{\mathtt{J}\star}) \in \mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}\star})$ and

$$r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star) \leq 0 \quad \forall \mathbf{x}_q \in \mathcal{P}_q(\mathbf{p}_{-q}^{\mathtt{J}\star}), \tag{3.46}$$

where $r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star)$ denotes the directional derivative of the function r_q^s at $\mathbf{x}_q^\star$ along the direction $\mathbf{x}_q - \mathbf{x}_q^\star$.

In words, a B-QGNE is a stationary solution of the GNEP, where the stationary concept is based on the directional derivative. Of course the B-QGNE can be defined also for the smooth game $\mathcal{G}^{\text{sm}}$; in such a case, the directional derivative $r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star)$ in (3.46) reduces to $\nabla_{\mathbf{x}_q} \widetilde{r}_q^s(\mathbf{x}_q^\star)^T(\mathbf{x}_q - \mathbf{x}_q^\star)$, because $\widetilde{r}_q^s(\bullet)$ is differentiable; thus, in the following, we will refer to it just as QGNE of $\mathcal{G}^{\text{sm}}$. Note that the B-QGNE is an instance of the Quasi-Nash Equilibrium introduced recently in [84, 85] to deal with the nonconvexity of the players' optimization problems.

It is worth mentioning that since the players' optimization problems in $\mathcal{G}$ and $\mathcal{G}^{\mathrm{sm}}$ have polyhedral feasible sets, every GNE of $\mathcal{G}$ (resp. $\mathcal{G}^{\mathrm{sm}}$) is a B-QGNE (resp. QGNE), but the converse is not necessarily true.

As already observed for the GNE, the subclass of quasi-solutions of $\mathcal{G}$ of practical interest are those associated with the strategy profiles belonging to the set $\mathcal{P}$. We can capture this feature introducing the concept of *restricted* B-QGNE of $\mathcal{G}$, which are B-QGNE "over the set $\mathcal{P}$"; hence, ignoring those feasible tuples of (3.41) that yield zero secrecy rates.

Definition 3.3 (Restricted B-QGNE)**.** A strategy $\mathbf{x}^\star \triangleq (\mathbf{x}_q^\star \triangleq (p_q^\star, \mathbf{p}_q^{\mathtt{J}\star}))_{q=1}^Q$ is a *restricted* B-QGNE of $\mathcal{G}$ if the following holds for all $q = 1, \ldots, Q$: $(p_q^\star, \mathbf{p}_q^{\mathtt{J}\star}) \in \mathcal{P}_q^{\mathtt{sm}}(\mathbf{p}_{-q}^{\mathtt{J}\star})$ and

$$r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star) \leq 0, \quad \forall \mathbf{x}_q \in \mathcal{P}_q^{\mathtt{sm}}(\mathbf{p}_{-q}^{\mathtt{J}\star}),$$

with $\mathcal{P}_q^{\mathtt{sm}}(\mathbf{p}_{-q}^{\mathtt{J}\star})$ defined in (3.44).

A natural question now is whether the connection between the GNE of $\mathcal{G}$ and $\mathcal{G}^{\mathtt{sm}}$ as stated in Proposition 3.4 is somehow preserved also in terms of quasi-equilibria (which in principle is not guaranteed). The answer is stated next.

Proposition 3.5. Given $\mathcal{G}$ and $\mathcal{G}^{\mathtt{sm}}$, the following hold.

(a) A B-QGNE of $\mathcal{G}$ always exists;

(b) A QGNE of $\mathcal{G}^{\mathtt{sm}}$ (resp. restricted B-QGNE of $\mathcal{G}$) exists provided that $\mathcal{P} \neq \emptyset$;

(c) $[\mathcal{G} \to \mathcal{G}^{\mathtt{sm}}]$: If $\mathbf{x}^\star$ is a (restricted) B-QGNE of $\mathcal{G}$ satisfying (3.43) for all $q = 1, \ldots, Q$, then $\mathbf{x}^\star$ is a QGNE of $\mathcal{G}^{\mathtt{sm}}$;

(d) $[\mathcal{G}^{\mathtt{sm}} \to \mathcal{G}]$: If $\mathbf{x}^\star$ is a QGNE of $\mathcal{G}^{\mathtt{sm}}$, then $\mathbf{x}^\star$ is a restricted B-QGNE of $\mathcal{G}$.

Proof. (*a*) It follows from Proposition 3.4(a) that a GNE of $\mathcal{G}$ always exists; since the players' optimization problems in $\mathcal{G}$ have polyhedral sets, every GNE of $\mathcal{G}$ is also a B-QGNE. Therefore a B-QGNE of $\mathcal{G}$ exists.

(b) It follows from Proposition 3.4(b) and similar arguments as in the proof of (a).

(c) This proof is based on the following fact, regarding the directional derivative of the plus function [13, Eq. 2.124]:

$$\begin{aligned} r_q^{s\prime}(\mathbf{z}; \mathbf{y}-\mathbf{z}) &= [\max(0, \widetilde{r}_q^s(\bullet))]'(\mathbf{z}; \mathbf{y}-\mathbf{z}) \\ &= \begin{cases} \max\left(0, \nabla_{\mathbf{x}_q} \widetilde{r}_q^s(\mathbf{z})^T(\mathbf{y}-\mathbf{z})\right), & \text{if } \widetilde{r}_q^s(\mathbf{z}) = 0 \\ \nabla_{\mathbf{x}_q} \widetilde{r}_q^s(\mathbf{z})^T(\mathbf{y}-\mathbf{z}), & \text{if } \widetilde{r}_q^s(\mathbf{z}) > 0. \end{cases} \end{aligned} \tag{3.47}$$

Let $\mathbf{x}^\star \triangleq \left(\mathbf{x}_q^\star\right)_{q=1}^Q$ be a (restricted) B-QGNE of $\mathcal{G}$ satisfying (3.43) for all $q = 1, \ldots, Q$, with $\mathbf{x}_q^\star \triangleq \left(\mathbf{p}_q^\star, \mathbf{p}_q^{\mathsf{J}\star}\right)$. Then, for every q, based on (3.47), consider the following two cases:

Case I: $\widetilde{r}_q^s(\mathbf{x}_q^\star) > 0$. For all $\mathbf{x}_q \in \mathcal{P}_q^{\mathsf{sm}}(\mathbf{p}_{-q}^{\mathsf{J}\star})$ we have

$$0 \geq r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star) = \nabla_{\mathbf{x}_q} \widetilde{r}_q^s(\mathbf{x}_q^\star)^T \left(\mathbf{x}_q - \mathbf{x}_q^\star\right).$$

Case II: $\widetilde{r}_q^s(\mathbf{x}_q^\star) = 0$. For all $\mathbf{x}_q \in \mathcal{P}_q^{\mathsf{sm}}(\mathbf{p}_{-q}^{\mathsf{J}\star})$ we have

$$0 \geq r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star) \geq \nabla_{\mathbf{x}_q} \widetilde{r}_q^s(\mathbf{x}_q^\star)^T \left(\mathbf{x}_q - \mathbf{x}_q^\star\right).$$

The desired result follows readily from the above two cases.

(d) Let $\mathbf{x}^\star \triangleq \left(\mathbf{x}_q^\star\right)_{q=1}^Q$ be a QGNE of $\mathcal{G}^{\mathsf{sm}}$, with $\mathbf{x}_q^\star \triangleq \left(\mathbf{p}_q^\star, \mathbf{p}_q^{\mathsf{J}\star}\right)$. Consider an arbitrary q. Clearly, $\mathbf{x}_q^\star \in \mathcal{P}_q^{\mathsf{sm}}(\mathbf{p}_{-q}^{\mathsf{J}\star})$. Based on (3.47), consider the following two cases:

Case I: $\widetilde{r}_q^s(\mathbf{x}_q^\star) > 0$. For all $\mathbf{x}_q \in \mathcal{P}_q^{\mathsf{sm}}(\mathbf{p}_{-q}^{\mathsf{J}\star})$ we have

$$r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star) = \nabla_{\mathbf{x}_q} \widetilde{r}_q^s(\mathbf{x}_q^\star)^T \left(\mathbf{x}_q - \mathbf{x}_q^\star\right) \leq 0$$

Case II: $\widetilde{r}_q^s(\mathbf{x}_q^\star) = 0$. For all $\mathbf{x}_q \in \mathcal{P}_q^{\mathsf{sm}}(\mathbf{p}_{-q}^{\mathsf{J}\star})$ we have

$$r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star) = \max\left(0, \nabla_{\mathbf{x}_q} \widetilde{r}_q^s(\mathbf{x}_q^\star)^T \left(\mathbf{x}_q - \mathbf{x}_q^\star\right)\right) = 0$$

The desired result comes readily from the above two cases. □

The above proposition paves the way to the design of numerical methods to compute a B-QGNE of $\mathcal{G}$. Indeed, according to statement (d), one can compute a B-QGNE of $\mathcal{G}$ (which is a solutions of practical interest) via a QGNE of $\mathcal{G}^{\mathtt{sm}}$. This last task is addressed in the next subsection.

3.5.1.4 Algorithmic Design

With the goal of computing a QGNE of $\mathcal{G}^{\mathtt{sm}}$ in mind, we capitalize on the potential structure of $\mathcal{G}^{\mathtt{sm}}$ and construct the following multiplayer linearly constrained optimization problem:

$$(\mathrm{P}): \quad \begin{array}{ll} \underset{(\mathbf{p},\mathbf{p}^{\mathrm{J}})}{\text{maximize}} & r(\mathbf{p},\mathbf{p}^{\mathrm{J}}) \triangleq \sum_{q=1}^{Q} \widetilde{r_q^s}(\mathbf{x}_q) \\ \text{subject to} & \left(\mathbf{p},\mathbf{p}^{\mathrm{J}}\right) \in \mathcal{P} \end{array} \tag{3.48}$$

The above nonconcave maximization problem is smooth, thus the standard definition of stationary solutions is applicable.

The connection between the social problem (P) in (3.48) and the games $\mathcal{G}$ and $\mathcal{G}^{\mathtt{sm}}$ is given in the next proposition.

Proposition 3.6. Given $\mathcal{G}^{\mathtt{sm}}$, and the social problem (P) in (3.48), the following hold.

(a) If $\mathcal{P} \neq \emptyset$, then (P) has an optimal solution;

(b) $[(P) \to \mathcal{G}^{\mathtt{sm}}]$: If $\mathbf{x}^\star$ is an optimal solution of (P), then $\mathbf{x}^\star$ is a GNE of $\mathcal{G}^{\mathtt{sm}}$;

(c) $[(P) \to \mathcal{G}^{\mathtt{sm}}]$: If $\mathbf{x}^\star$ is a stationary solution of (P), then $\mathbf{x}^\star$ is a QGNE of $\mathcal{G}^{\mathtt{sm}}$;

(d) $[\mathcal{G}^{\mathtt{sm}} \to (P)]$: If $\mathbf{x}^\star$ is a QGNE of $\mathcal{G}^{\mathtt{sm}}$ and there exists common multipliers of the shared constraints $\sum_{r=1}^{Q} p_{jr}^{\mathrm{J}} \leq P_j^{\mathrm{J}} \quad j = 1, \ldots, J$ for all players, then $\mathbf{x}^\star$ is a stationary solution of (P).

Proof. (a) This is clear.

(b) Since the game $\mathcal{G}^{\mathtt{sm}}$ is of the potential type, the desired result follows immediately.

(c) For notational simplicity, generalizing (3.48), consider the following abstract optimization problem with separable, differentiable, objective function:

$$\left.\begin{array}{ll}\underset{\mathbf{x}\triangleq(\mathbf{x}_q)_{q=1}^Q}{\text{maximize}} & \displaystyle\sum_{q=1}^{Q}\psi_q(\mathbf{x}_q) \\ \text{subject to:} & \mathbf{x}\in X\triangleq\prod_{q=1}^{Q}X_q, \quad \displaystyle\sum_{q=1}^{Q}\mathbf{A}_q\mathbf{x}_q\leq\mathbf{b},\end{array}\right\}\triangleq\widehat{\Xi} \tag{3.49}$$

where each $\mathbf{x}_q\in\mathbb{R}^{n_q}$, $X_q\subset\mathbb{R}^{n_q}$ is a polyhedron, $\mathbf{A}_q$ is a matrix of order $\mathbb{R}^{m\times n_q}$, $\mathbf{b}$ is a m-dimensional vector, and $\widehat{\Xi}$ denotes the feasible set of (3.49). Let $\mathbf{x}^\star\triangleq\left(\mathbf{x}_q^\star\right)_{q=1}^Q$ be a stationary solution of (3.49). By definition, this means that $\mathbf{x}^\star\in\widehat{\Xi}$ and $\sum_{q=1}^{Q}\left(\mathbf{x}_q-\mathbf{x}_q^\star\right)^T\nabla_{\mathbf{x}_q}\psi_q(\mathbf{x}_q^\star)\leq 0$, for all $\mathbf{x}\in\widehat{\Xi}$. In particular, for every $\mathbf{x}_q\in X_q$ such that: $\mathbf{A}_q\mathbf{x}_q+\sum_{r\neq q}\mathbf{A}_r\,\mathbf{x}_r^\star\leq\mathbf{b}$, we have $\left(\mathbf{x}_q-\mathbf{x}_q^\star\right)^T\nabla\psi_q(\mathbf{x}_q^\star)\leq 0$, which establishes that $\mathbf{x}_q^\star$ is a stationary solution of the optimization problem:

$$\underset{\mathbf{x}_q\in X_q\,:\,\mathbf{A}_q\mathbf{x}_q+\sum_{r\neq q}\mathbf{A}_r\,\mathbf{x}_r^\star\leq\mathbf{b}}{\text{maximize}}\quad\psi_q(\mathbf{x}_q)$$

Hence $\mathbf{x}^\star$ is a QGNE of the game where each player q, anticipating $(\mathbf{x}_r)_{r\neq q}$, solves the optimization problem:

$$\underset{\mathbf{x}_q\in X_q\,:\,\mathbf{A}_q\mathbf{x}_q+\sum_{r\neq q}\mathbf{A}_r\,\mathbf{x}_r\leq\mathbf{b}}{\text{maximize}}\quad\psi_q(\mathbf{x}_q)$$

Thus, the desired conclusion follows easily by applying this general result to the secrecy rate game $\mathcal{G}^{\mathtt{sm}}$.

(d) This result follows readily under the common multipliers assumption. $\square$

Figure 3.2 summarizes the main results and relationship between $\mathcal{G}$, $\mathcal{G}^{\mathtt{sm}}$ and the social problem (P), as stated in Propositions 3.4, 3.5 and 3.6.

Based on Proposition 3.6, one can now design distributed algorithms for computing a (Q)GNE of $\mathcal{G}^{\mathtt{sm}}$ (and thus a restricted B-QGNE of the original game $\mathcal{G}$): it is sufficient to solve the social problem (P). Such a problem is nonconcave; stationary solutions however can be computed efficiently observing that (P) is an instance of the DC program (3.1), under the following

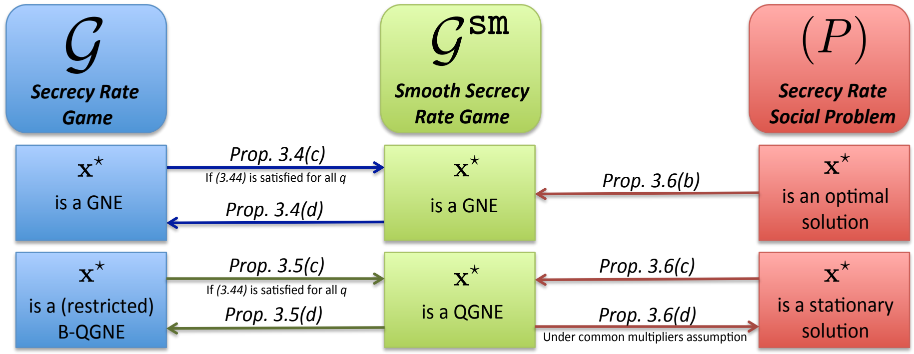

Figure 3.2: Connections between the games $\mathcal{G}$, $\mathcal{G}^{\mathtt{sm}}$ and the social problem (P).

identifications:

$$f_q\left(p_q,\mathbf{p}_q^{\mathtt{J}}\right) \triangleq -\log\left(\sigma^2 + H_{qq}^{\mathtt{SD}} p_q + \sum_{j=1}^{J} H_{jq}^{\mathtt{JD}}\, p_{jq}^{\mathtt{J}}\right) - \log\left(\sigma^2 + \sum_{j=1}^{J} H_{je}^{\mathtt{JE}}\, p_{jq}^{\mathtt{J}}\right)$$

$$g_q\left(p_q,\mathbf{p}_q^{\mathtt{J}}\right) \triangleq -\log\left(\sigma^2 + H_{qe}^{\mathtt{SE}} p_q + \sum_{j=1}^{J} H_{je}^{\mathtt{JE}}\, p_{jq}^{\mathtt{J}}\right) - \log\left(\sigma^2 + \sum_{j=1}^{J} H_{jq}^{\mathtt{JD}}\, p_{jq}^{\mathtt{J}}\right).$$

Then, we can compute efficiently a stationary solution of (P) using any of the distributed algorithms introduced in Section 3.4. For instance, Algorithm 3.1 based on a dual decomposition loop (cf. Algorithm 3.3) specialized to (P) is given in Algorithm 3.4, where in (3.51), we let

$$\mathcal{X}_q \triangleq \left\{\mathbf{x}_q \triangleq (p_q,\mathbf{p}_q^{\mathtt{J}}) \geq \mathbf{0} \;:\; p_q \leq P_q \text{ and } (3.43) \text{ holds}\right\}$$

and

$$\widetilde{\theta}_q(\mathbf{x}_q;\mathbf{x}_q^{\nu}) \triangleq f_q\left(\mathbf{x}_q\right) - \frac{\partial g_q(\mathbf{x}_q^{\nu})}{\partial p_q}\, p_q - \sum_{j=1}^{J} \frac{\partial g_q(\mathbf{x}_q^{\nu})}{\partial p_{jq}^{\mathtt{J}}}\, p_{jq}^{\mathtt{J}} + \frac{\tau_q}{2}\|\mathbf{x}_q - \mathbf{x}_q^{\nu}\|^2, \quad (3.50)$$

with

$$\frac{\partial g_q(\mathbf{x}_q)}{\partial p_q} \triangleq -\frac{H_{qe}^{\mathtt{SE}}}{\sigma^2 + H_{qe}^{\mathtt{SE}} p_q + \sum_{j=1}^{J} H_{je}^{\mathtt{JE}}\, p_{jq}^{\mathtt{J}}}$$

$$\frac{\partial g_q(\mathbf{x}_q)}{\partial p_{jq}^{\mathtt{J}}} \triangleq -\frac{H_{je}^{\mathtt{JE}}}{\sigma^2 + H_{qe}^{\mathtt{SE}} p_q + \sum_{j=1}^{J} H_{je}^{\mathtt{JE}}\, p_{jq}^{\mathtt{J}}} - \frac{H_{jq}^{\mathtt{JD}}}{\sigma^2 + \sum_{j=1}^{J} H_{jq}^{\mathtt{JD}}\, p_{jq}^{\mathtt{J}}}, \quad j = 1,\ldots,J.$$

Algorithm 3.4: DC-based Algorithm for (P)

Data: $\boldsymbol{\tau} \triangleq (\tau_q)_{q=1}^{Q} \geq \mathbf{0}, \{\gamma^\nu\} > 0$, $\{\alpha^t\} > 0$ and $\mathbf{x}^0 \triangleq (\mathbf{p}^0, \mathbf{p}^{\mathtt{J},0}) \in \mathcal{P}$. Set $\nu = 0$.

(S.1): If $\mathbf{x}^\nu \triangleq (\mathbf{p}^\nu, \mathbf{p}^{\mathtt{J},\nu})$ satisfies a termination criterion, STOP.

(S.2a): Choose $\boldsymbol{\lambda}^0 \triangleq (\lambda_j^0)_{j=1}^{J} \geq \mathbf{0}$. Set $t = 0$.

(S.2b): If $\boldsymbol{\lambda}^t \triangleq (\lambda_j^t)_{j=1}^{J}$ satisfies a termination criterion, set $\widehat{\mathbf{x}}(\mathbf{x}^\nu) = (p_q^{\nu,t}, \mathbf{p}_q^{\mathtt{J},\nu,t})_{q=1}^{Q}$ and go to (S.3).

(S.2c): The legitimate users $q = 1, \ldots, Q$ compute in parallel $(p_q^{\nu,t}, \mathbf{p}_q^{\mathtt{J},\nu,t})$ given by

$$(p_q^{\nu,t}, \mathbf{p}_q^{\mathtt{J},\nu,t}) \triangleq \underset{\mathbf{x}_q \triangleq (p_q, \mathbf{p}_q^{\mathtt{J}}) \in \mathcal{X}_q}{\operatorname{argmin}} \left\{ \widetilde{\theta}_q(\mathbf{x}_q; \mathbf{x}_q^\nu) + \boldsymbol{\lambda}^{tT} \mathbf{p}_q^{\mathtt{J}} \right\}. \tag{3.51}$$

(S.2d): Update $\boldsymbol{\lambda} \triangleq (\lambda_j)_{j=1}^{J}$: for all $j = 1, \ldots, J$,

$$\lambda_j^{t+1} \triangleq \left[\lambda_j^t + \alpha^t \left(\sum_{q=1}^{Q} p_{jq}^{\mathtt{J},\nu,t} - P_j^{\mathtt{J}} \right) \right]^+. \tag{3.52}$$

(S.2e): Set $t \leftarrow t + 1$ and go back to (S.2b).

(S.3): Set $\mathbf{x}^{\nu+1} = \mathbf{x}^\nu + \gamma^\nu \left(\widehat{\mathbf{x}}(\mathbf{x}^\nu) - \mathbf{x}^\nu \right)$.

(S.4): $\nu \leftarrow \nu + 1$ and go to (S.1).

If the sequences $\{\gamma^\nu\} > 0$ and $\{\alpha^t\} > 0$ are chosen according to one of the rules stated in Theorems 3.1 and 3.4, respectively, Algorithm 3.4 converges to a stationary solution of the social problem (P) (in the sense of Theorem 3.1), and thus to a QGNE of $\mathcal{G}^{\mathtt{sm}}$ [cf. Proposition 3.6(c)], which is also a restricted B-QGNE of $\mathcal{G}$ [cf. Proposition 3.5(d)].

Remark 3.3 (On the implementation of Algorithm 3.4)**.** Once the CSI is available at the users' sides, Algorithm 3.4 can be implemented in a distributed way, with limited signaling only between the legitimate users and the friendly jammers (no communications among the users or the eavesdropper is required). Indeed, in the inner loop of the algorithm, given the current value of the price $\boldsymbol{\lambda}$, all the users simultaneously update the power profiles $(p_q, \mathbf{p}_q^{\mathtt{J}})$ solving locally a strongly convex optimization problem. Then, they communicate to the friendly jammers the amount of power they need resulting from the optimization. Given the power requests from the users, the jammers update in parallel and independently their price λ_j performing an inexpensive scalar projection [cf. (4.38)], and then broadcast the new price value to the legitimate users. We remark that the proposed scheme requires the same CSI and communication overhead than CJ approaches proposed in the literature (see, e.g., [36, 104]).

Remark 3.4 (More general formulation)**.** For the sake of simplicity, in the previous sections, we assumed uniform power allocation over the spectrum (still to optimize) from the users and the friendly jammers. We remark however that game $\mathcal{G}$ in (3.41) [game $\mathcal{G}^{\mathtt{sm}}$ in (3.44) and problem (P) in (3.48)] can be generalized to the case of nonuniform power allocations (over flat-fading channels) and the proposed algorithms extended accordingly.

3.5.1.5 Numerical Results

In this subsection, we present some experiments validating our theoretical findings. We compare our Algorithm 3.4 with centralized and applicable decentralized schemes existing in the literature (adapted to our formulation).

System Setup. All the experiments are obtained in the following setting, unless stated otherwise. All the users and jammers have the same power budget, i.e. $P_q = P_j^{\mathtt{J}} = P$, and we set $\mathtt{snr} = P/\sigma^2 = 10$dB. The position of the users, jammers, and eavesdropper are randomly generated within a square area; the channel gains $H_{qq}^{\mathtt{SD}}$, $H_{qe}^{\mathtt{SE}}$, $H_{jq}^{\mathtt{JD}}$ and $H_{je}^{\mathtt{JE}}$ are Rayleigh distributed with mean equal to one and normalized by the (square) distance between the transmitter and the receiver; our results are collected only for the channel realizations satisfying condition (3.43). When present, there are $J = \lfloor Q/2 \rfloor$ jammers. Algorithm 3.4 is initialized by choosing the zero power allocation, and it is terminated when the absolute value of the difference of the System Secrecy Rate (SSR) in two consecutive iterations becomes smaller than $1e$-5. Similarly, the inner loop is terminated when the difference of the norm of the prices in two consecutive rounds is less than $1e$-2.

Example 3.1. Comparison with Decentralized Schemes. In Figure 3.3(a), we plot the average SSR (taken over 50 independent channel realizations) versus the number Q of legitimate users achieved by our Algorithm 3.4 (blue-line curves) and by solving the SSR maximization game while assuming i) uniform power allocation for the jammers (black-line curves); or

ii) no friendly jammers available (red-line curve). In Figure 3.3(b) we plot the average SSR versus the $\mathtt{snr}$ for the case of 10 main users; the rest of the setting is as in Figure 3.3(a). The figures show that the proposed approach yields much higher SSR than that achievable by the other schemes, and the gain becomes more significant as the number of users or the $\mathtt{snr}$ increases.

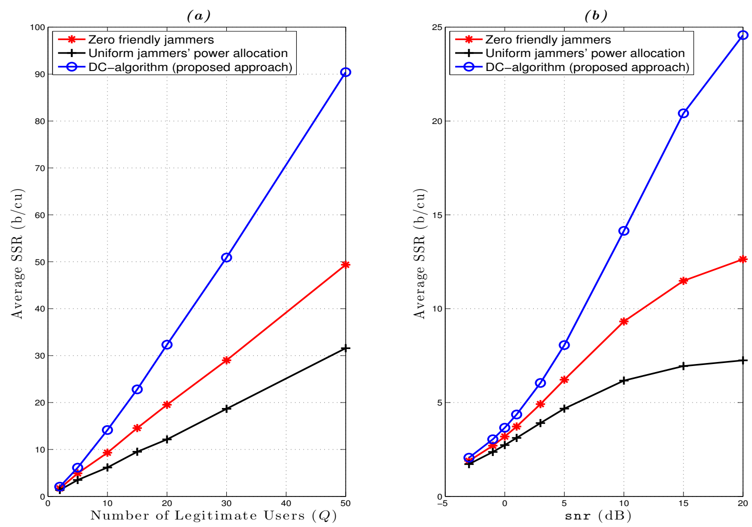

Figure 3.3: Average system secrecy rate (SSR) versus (a) number Q of legitimate users; (b) $\mathtt{snr}$, for $Q = 10$.

Example 3.2. Comparison with Centralized Schemes. Since Algorithm 3.4 converges to a stationary solution of the social problem (3.48), it is natural to compare our scheme with available centralized methods attempting to compute locally or globally optimal solutions (but without rigorously verifying their optimality). More specifically, we consider two schemes: i) the NEOS server [23] based on MINOS solver and PSwarm; and ii) the standard centralized SCA algorithm for DC programs (see, e.g., [5, 90]). Even thought these algorithms are computationally very demanding and not implementable in a distributed network, they represent a good benchmark to test our distributed algorithm. In Figure 3.4(a) we plot the probability that the SSR exceeds a given value SSR versus SSR, whereas in Figure 3.4(b) we report the average SSR versus the number of legitimate users. All the curves are computed running 100 independent experiments. The figures show that our Algorithm 3.4 outperforms MINOS, and quite surprisingly it has the same performance of PSwarm and SCA schemes. This means that, at least for the experiments we simulated, Algorithm 3.4 provides *in a distributed way* solutions of (3.48) that are very close to those obtained by centralized methods.

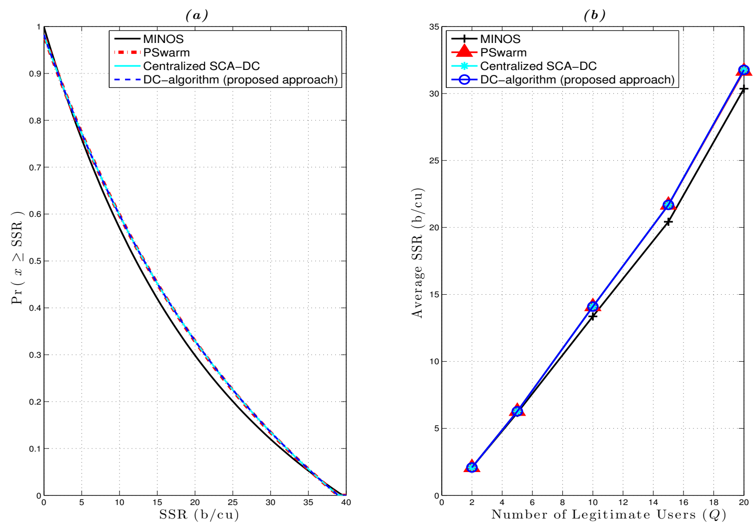

Figure 3.4: Comparison between our (distributed) Algorithm 3.4 and centralized approaches.

Example 3.3. Convergence Speed. Figure 3.5 shows the average number of inner and outer iterations required by Algorithm 3.4 to converge versus the number of legitimate users, when Rules 1 and 2 [c.f. (3.21) and (3.22)] are used for the step-size γ^{ν}. The parameters in the two rules are set as $\beta_1 = 1e$-6, $\beta_2 = 1e$-4, and $\epsilon = 1e$-5. This average has been taken over 50 independent channel realizations. Notice that, when Rule 2 is used, the number of outer iterations is greatly reduced in comparison to that obtained performing Rule 1. Note also that, the average number of inner iterations per outer iteration is slightly increased in the former case; however, overall, Algorithm 3.4 with Rule 2 seems to be faster than the same algorithm based on Rule 1. As far as the quality of the achieved SSR is concern, we observed results consistent to Figure 3.4.

3.5.2 Case Study 2: Design of CR MIMO Systems

Current wireless communication systems are characterized by *fixed* spectrum assignment policies. However, such fixed policies are known to be very inefficient since the utilization of the licensed bandwidth is highly variant. This

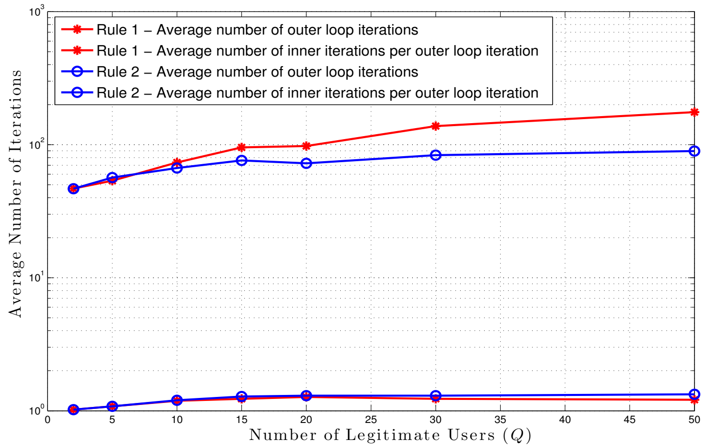

Figure 3.5: Average number of outer iterations and inner iterations per outer iteration in Algorithm 3.4 for different choices of diminishing step-size rules.

issue, coupled with the recent increasing demand of wireless services, has originated a problem of scarcity of the radio spectrum. Based on the dynamic nature of the spectrum' usage, Cognitive Radio (CR) has been proposed as a solution to the cited scarcity problem. Roughly speaking, the main idea behind Cognitive Radio is to share the utilization of the spectrum between licensed and unlicensed spectrum holders, by providing the network entities with the ability to adapt their transmission parameters dynamically in response to the current state of the system [99]; we refer the interested reader to [38, 127, 99] and the references therein for more information about this topic. The CR paradigm introduces a hierarchical structure in the network, where the licensed spectrum holders, also known as Primary Users (PUs), allow the so-called Secondary Users (SUs), i.e. the unlicensed spectrum holders, to access the licensed spectrum dynamically with the condition of not degrading their performance.

In the context described above, we consider the sum-rate maximization of CR MIMO systems, subject to coupling interference constraints. Special cases of such a nonconvex problem have been widely studied in the literature. The analysis is mainly limited to local interference constraints (see [93, 124] and references therein), with the exception of [51, 126] where coupling constraints are considered. However the theoretical convergence of algorithms in [51, 126] is up to date an *open problem*. Since the general optimization problem is an instance of DC programs with shared constraints, we can apply the framework developed in the first part of this chapter and obtain for the first time a class of distributed (primal/dual-based) algorithms with provable convergence.

3.5.2.1 System Model and Problem Formulation

Consider a hierarchical MIMO CR system composed of P primary users sharing the licensed spectrum with I secondary users; the network of SUs is modeled as a I-user vector Gaussian interference channel. Each SU i is equipped with n_{T_i} and n_{R_i} transmit and receive antennas, respectively, and PUs may have multiple antennas. Let $\mathbf{H}_{ji} \in \mathbb{C}^{n_{R_j} \times n_{T_i}}$ (resp. $\mathbf{G}_{pi}$) be the cross-channel matrix between the secondary transmitter i and the secondary receiver j (resp. primary receiver p). Under basic information theoretical assumptions, the transmission rate of SU i is

$$r_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) \triangleq \log\det\left(\mathbf{I} + \mathbf{H}_{ii}^H \mathbf{R}_i(\mathbf{Q}_{-i})^{-1} \mathbf{H}_{ii} \mathbf{Q}_i\right) \tag{3.53}$$

where $\mathbf{Q}_i$ is the transmit covariance matrix of SU i (to be optimized), $\mathbf{Q}_{-i} \triangleq (\mathbf{Q}_j)_{j\neq i}$, $\mathbf{R}_i(\mathbf{Q}_{-i}) \triangleq \mathbf{R}_{n_i} + \sum_{j\neq i} \mathbf{H}_{ij}\mathbf{Q}_j\mathbf{H}_{ij}^H$ with $\mathbf{R}_{n_i} \succ \mathbf{0}$ being the covariance matrix of the noise plus the interference generated by the active PUs. Each SU i is subject to the following local constraints

$$\mathcal{Q}_i \triangleq \left\{ \mathbf{Q}_i \succeq \mathbf{0} : \ \mathrm{tr}(\mathbf{Q}_i) \leq P_i^{\mathrm{tot}}, \ \mathbf{Q}_i \in \mathcal{Z}_i \ \right\}, \tag{3.54}$$

where P_i^{tot} is the total transmit power and $\mathcal{Z}_i \subseteq \mathbb{C}^{n_i \times n_i}$ is an abstract closed and convex set suitable to accommodate (possibly) additional local constraints, such as: i) *null constraints* $\mathbf{U}_i^H\mathbf{Q}_i = \mathbf{0}$, with $\mathbf{U}_i$ being $n_{T_i} \times r_{U_i}$ (with $r_{U_i} < n_{T_i}$), which prevents SUs to transmit along some prescribed "directions" (the columns of $\mathbf{U}_i$); and ii) *soft and peak power shaping constraints* in the form of $\mathrm{tr}(\mathbf{T}_i^H\mathbf{Q}_i\mathbf{T}_i) \leq I_i^{\mathrm{ave}}$ and $\lambda^{\max}(\mathbf{F}_i^H\mathbf{Q}_i\mathbf{F}_i) \leq I_i^{\mathrm{peak}}$, which limits to $I_i^{\mathrm{ave}} > 0$ and $I_i^{\mathrm{peak}} > 0$ the total average and peak average power allowed to be radiated along the range space of matrices $\mathbf{T}_i \in \mathbb{C}^{n_{T_i} \times n_{G_i}}$ and $\mathbf{F}_i \in \mathbb{C}^{n_{T_i} \times n_{F_i}}$, respectively. Adopting the spectrum underlay architecture, the interference power at the primary receivers is regulated by imposing global interference constraints to the SUs, in the form of $\sum_i \mathrm{tr}(\mathbf{G}_{pi}^H\mathbf{Q}_i\mathbf{G}_{pi}) \leq I_p^{\mathrm{tot}}$ for all $p = 1, \ldots, P$, where $I_p^{\mathrm{tot}} > 0$ is the interference threshold imposed by PU p.

The design of the CR system can be formulated as

$$\begin{array}{ll} \underset{\mathbf{Q}_1,\ldots,\mathbf{Q}_I}{\text{maximize}} & \theta(\mathbf{Q}) \triangleq \sum_{i=1}^{I} r_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) \\ \text{subject to} & \mathbf{Q}_i \in \mathcal{Q}_i, \quad i = 1, \ldots, I, \\ & \sum_{i=1}^{I} \operatorname{tr}(\mathbf{G}_{pi}^H \mathbf{Q}_i \mathbf{G}_{pi}) \leq I_p^{\text{tot}}, \quad p = 1, \ldots, P. \end{array} \tag{3.55}$$

We remark that special cases of the nonconvex sum-rate maximization problem (3.55) have already been studied in the literature, e.g., in [51, 126]. However the theoretical convergence of current algorithms is up to date an open problem. Since (3.55) is an instance of (3.1), we can capitalize on the framework proposed in this chapter and obtain readily a class of distributed algorithms with *provable convergence*.

3.5.2.2 DC-based Decomposition Algorithms

We cast first (3.55) into (3.1) (in the maximization form). Exploring the DC-structure of the rates $r_i(\mathbf{Q}_i, \mathbf{Q}_{-i})$, the sum-rate $\theta(\mathbf{Q})$ can indeed be rewritten as the sum of a concave and convex function, namely:

$$\theta(\mathbf{Q}) = \sum_{i=1}^{I} \left(f_i(\mathbf{Q}) - g_i(\mathbf{Q}) \right),$$

where

$$\begin{aligned} f_i(\mathbf{Q}) &\triangleq \log\det\left(\mathbf{R}_{n_i} + \textstyle\sum_{j=1} \mathbf{H}_{ij}\mathbf{Q}_j\mathbf{H}_{ij}^H\right) \\ g_i(\mathbf{Q}_{-i}) &\triangleq \log\det\left(\mathbf{R}_{n_i} + \textstyle\sum_{j\neq i} \mathbf{H}_{ij}\mathbf{Q}_j\mathbf{H}_{ij}^H\right). \end{aligned}$$

We can now use the machinery developed in Section 3.3. It is not difficult to show that the approximation function $\widetilde{\theta}(\mathbf{x}; \mathbf{x}^\nu)$ in (3.5) (to be maximized) becomes (up to a constant term)

$$\widetilde{\theta}(\mathbf{Q}; \mathbf{Q}^\nu) \triangleq \sum_{i=1}^{I} \widetilde{\theta}_i(\mathbf{Q}; \mathbf{Q}^\nu),$$

with

$$\widetilde{\theta}_i(\mathbf{Q};\mathbf{Q}^\nu) \triangleq r_i(\mathbf{Q}_i,\mathbf{Q}_{-i}^\nu) - \langle \mathbf{\Pi}_i(\mathbf{Q}^\nu), \mathbf{Q}_i - \mathbf{Q}_i^\nu \rangle - \frac{\tau_i}{2}\left\|\mathbf{Q}_i - \mathbf{Q}_i^\nu\right\|_F^2 \tag{3.56}$$

where $\mathbf{Q}^\nu \triangleq (\mathbf{Q}_i^\nu)_{i=1}^I$ with each $\mathbf{Q}_i^\nu \succeq \mathbf{0}$, $\langle \mathbf{A},\mathbf{B}\rangle \triangleq \mathrm{Re}\left\{\mathrm{tr}(\mathbf{A}^H\mathbf{B})\right\}$, $\|\bullet\|_F$ is the Frobenius norm, and

$$\mathbf{\Pi}_i(\mathbf{Q}^\nu) \triangleq \sum_{j\in\mathcal{N}_i} \mathbf{H}_{ji}^H\,\widetilde{\mathbf{R}}_j(\mathbf{Q}_{-j}^\nu)\,\mathbf{H}_{ji}, \tag{3.57}$$

with $\mathcal{N}_i$ denoting the set of neighbors of user i (i.e., the set of users j's which user i interferers with), and

$$\widetilde{\mathbf{R}}_j(\mathbf{Q}_{-j}^\nu) \triangleq \mathbf{R}_j(\mathbf{Q}_{-j}^\nu)^{-1} - (\mathbf{R}_j(\mathbf{Q}_{-j}^\nu) + \mathbf{H}_{jj}\mathbf{Q}_j^\nu\mathbf{H}_{jj}^H)^{-1}.$$

Therefore a stationary solution of (3.55) can be efficiently computed using any of the distributed algorithms introduced in Section 3.4; one just needs to replace $\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu)$ with $\widetilde{\theta}(\mathbf{Q};\mathbf{Q}^\nu)$ (and the minimization with the maximization). For instance, Algorithm 3.1 based on a dual decomposition loop (cf. Algorithm 3.3) can be written in the form of Algorithm 3.5. Convergence (in the sense of Theorem 3.1) is guaranteed if the step-size sequences $\{\gamma^\nu\}$ and $\{\alpha^n\} > 0$ are chosen according to one of the rules stated in Theorems 3.1 and 3.4, respectively.

Remark 3.5 (On the Implementation of Algorithm 3.5)**.** This algorithm is a double loop scheme in the sense described next.

Inner loop: In this loop, at every iteration t:

i) First, all SUs solve in *parallel* their strongly convex optimization problems (3.58), for fixed $\boldsymbol{\lambda}^t = (\lambda_p^t)_{p=1}^P$, resulting in the optimal solutions $(\mathbf{Q}_i^{\nu,t})_{i=1}^I$.

ii) Then, given the new interference levels $\sum_{i=1}^I \mathrm{tr}(\mathbf{G}_{pi}^H\mathbf{Q}_i^{\nu,t}\mathbf{G}_{pi})$, the prices $\boldsymbol{\lambda}$ are updated in parallel via (3.59), resulting in $\boldsymbol{\lambda}^{t+1} = (\lambda_p^{t+1})_{p=1}^P$.

The loop terminates when $\{\boldsymbol{\lambda}^t\}$ meets the termination criterion in (S.2b).

Outer loop: It consists in updating $\mathbf{Q}_i^\nu$'s according to (S.3).

Algorithm 3.5: DC-based Algorithm for (3.55).

Data: $\boldsymbol{\tau} \triangleq (\tau_i)_{i=1}^I \geq \mathbf{0}, \{\gamma^\nu\} > 0$, $\{\alpha^t\} > 0$ and $\mathbf{Q}_i^0 \in \mathcal{Q}_i$ for all i. Set $\nu = 0$.

(`S.1`): If $\mathbf{Q}^\nu \triangleq (\mathbf{Q}_i^\nu)_{i=1}^I$ satisfies a termination criterion, STOP.

(`S.2a`): Choose $\boldsymbol{\lambda}^0 \triangleq (\lambda_p^0)_{p=1}^P \geq \mathbf{0}$. Set $t = 0$.

(`S.2b`): If $\boldsymbol{\lambda}^t \triangleq (\lambda_p^t)_{p=1}^P$ satisfies a termination criterion, set $\widehat{\mathbf{Q}}(\mathbf{Q}^\nu) \triangleq (\mathbf{Q}_i^{\nu,t})_{i=1}^I$ and go to (`S.3`).

(`S.2c`): The SUs solve in parallel the following strongly convex optimization problems: for all $i = 1, \ldots, I$,

$$\mathbf{Q}_i^{\nu,t} \triangleq \underset{\mathbf{Q}_i \in \mathcal{Q}_i}{\text{argmax}} \left\{ \widetilde{\theta}_i(\mathbf{Q}_i; \mathbf{Q}^\nu) - \sum_{p=1}^P \lambda_p^t \operatorname{tr}(\mathbf{G}_{pi}^H \mathbf{Q}_i \mathbf{G}_{pi}) \right\}. \tag{3.58}$$

(`S.2d`): Update $\boldsymbol{\lambda} \triangleq (\lambda_p)_{p=1}^P$: for all $p = 1, \ldots, P$,

$$\lambda_p^{t+1} \triangleq \left[\lambda_p^t + \alpha^t \left(\sum_{i=1}^I \operatorname{tr}(\mathbf{G}_{pi}^H \mathbf{Q}_i^{\nu,t} \mathbf{G}_{pi}) - I_p^{\max} \right) \right]^+. \tag{3.59}$$

(`S.2e`): Set $t \leftarrow t + 1$ and go back to (`S.2b`).

(`S.3`): Set $\mathbf{Q}^{\nu+1} = \mathbf{Q}^\nu + \gamma^\nu \left(\widehat{\mathbf{Q}}(\mathbf{Q}^\nu) - \mathbf{Q}^\nu \right)$.

(`S.4`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

Communication overhead: The proposed algorithm is fairly distributed. Indeed, given the interference generated by the other users (the covariance matrix $\mathbf{R}_i(\mathbf{Q}_{-i})$, which can be locally measured), and the interference price $\mathbf{\Pi}_i(\mathbf{Q}^\nu)$, each SU can efficiently and *locally* compute the optimal covariance matrix $\mathbf{Q}_i^{\nu,t}$ by solving (3.58). Note that, for some specific structures of the feasible sets $\mathcal{Q}_i$ and channels (e.g., $\mathcal{Z}_i = \emptyset$, full-column rank channel matrices $\mathbf{H}_{ii}$, and $\tau_i = 0$), a solution of (3.58) is available in closed form (up to the multipliers associated with the power budget constraints) [51]. The estimation of the prices $\mathbf{\Pi}_i(\mathbf{Q}^\nu)$ requires some signaling exchange but only among nearby users. Interestingly, the pricing expression (3.57) as well as the signaling overhead necessary to compute it coincides with that of pricing schemes proposed in the literature to solve related problems [51, 94, 93].

The natural candidates for updating the prices in the inner loop are the primary users, after measuring *locally* the current overall interference given by $\sum_{i=1}^I \operatorname{tr}(\mathbf{G}_{pi}^H \mathbf{Q}_i^{\nu,t} \mathbf{G}_{pi})$ from the SUs. Note that this update is computationally inexpensive (it is a projection onto $\mathbb{R}_+$), it can be performed in parallel among PUs, and does not require any signaling exchange with the SUs. The new value of the prices is then broadcast to the SUs. In CR scenarios where the PUs cannot participate in the updating process, the SUs themselves can perform the price update, at the cost of more signaling, e.g., using consensus algorithms. Alternatively, if the primary receivers have a fixed geographical location, it might be possible to install some monitoring devices close to each primary receiver having the functionality of price computation and broadcasting.

As a final remark note that, since Algorithm 3.5 is a dual-based scheme, it is scalable with the number of SUs. However, for the same reasons, there might happen that the interference constraints are not satisfied during the intermediate iterations. This issue can be alleviated in practice by choosing a "large" $\boldsymbol{\lambda}^0$ as initial price. An alternative distributed scheme which does not suffer from this issue can be readily obtained using the primal-based decomposition approach introduced in Subsection 3.4.2.

3.6 Conclusion

In this chapter we proposed a novel decomposition framework to compute stationary solutions of nonconvex (possibly DC) sum-utility minimization problems with coupling convex constraints. We developed a class of (inexact) best-response-like algorithms, where all the users iteratively solve in *parallel* a suitably convexified version of the original DC program. To the best of our knowledge, this is the first set of *distributed* algorithms *with provable convergence* for multiuser DC programs with *coupling constraints*. Finally, we tested our methodology on two problems: i) a novel secrecy rate game, for which we developed algorithms to compute its QGNE based on a nontrivial DC reformulation; and ii) the sum-rate maximization problems over CR MIMO networks, for which we provided (for the first time) provable convergent distributed algorithms. Experiments show that our distributed algorithms reach performance comparable (and sometimes better) than centralized schemes.

As a future research topic, we suggest the extension of the results in this chapter to nonconvex constraints. Furthermore, we have also assumed the differentiability of both the objective functions and the constraints. There are diverse applications in this field that violate the aforementioned assumptions, hence this topic is worth to be explored.

Chapter 4

Maximization of the Sum of Max Functions and its Applications[1]

4.1 Introduction

Consider a resource allocation problem in a multiuser system composed of I users. Each user $i = 1, \ldots, I$ makes his decision on an n_i-dimensional real strategy vector $\mathbf{x}_i \in \mathbb{R}^{n_i}$ subject to some private and coupling constraints. Let $\mathbf{x} \triangleq (\mathbf{x}_i)_{i=1}^{I} \in \mathbb{R}^n$, where $n = \sum_{i=1}^{I} n_i$, be the strategy profile vector of all the users in the system; the vector $\mathbf{x}_{-i} \triangleq (\mathbf{x}_\ell)_{\ell \neq i}$ denotes the strategies of all the users except user i. The local constraints are given by the sets $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$ $i = 1, \ldots, I$; and, the n_s side constraints are represented by $\mathbf{s}(\mathbf{x}) \triangleq (s_k(\mathbf{x}))_{k=1}^{n_s} \leq \mathbf{0}$. The joint strategy set is denoted by $\mathcal{X} \triangleq \Pi_{i=1}^{I} \mathcal{X}_i$; and $\mathcal{X}_{-i} \triangleq \Pi_{\ell \neq i} \mathcal{X}_\ell$. We are interested in studying a sum-utility function of a particular kind, namely, the utility function θ_i of each user i is a continuous max function of the form:

$$\theta_i(\mathbf{x}) \triangleq \operatorname*{maximum}_{\boldsymbol{\lambda}_i \in \Lambda_i} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i) f_{i,j}(\mathbf{x}). \tag{4.1}$$

Hence, the system design is formulated as an optimization problem of the following kind, which for shortness, we call MSM (Maximization of the Sum of Max functions):

$$\begin{array}{lll} \operatorname*{maximize}_{\mathbf{x} \triangleq (\mathbf{x}_i)_{i=1}^{I}} & \theta(\mathbf{x}) \triangleq \displaystyle\sum_{i=1}^{I} \left(\operatorname*{maximum}_{\boldsymbol{\lambda}_i \in \Lambda_i} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i) f_{i,j}(\mathbf{x}) \right) & \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i \quad \forall\, i = 1, \ldots, I & \textit{(private constraints)} \\ & \mathbf{s}(\mathbf{x}) \triangleq (s_k(\mathbf{x}))_{k=1}^{n_s} \leq \mathbf{0} & \textit{(coupling constraints).} \end{array} \tag{4.2}$$

[1] This chapter is adapted from a manuscript in preparation for submission (Co-authors: Jong-Shi Pang, Gesualdo Scutari and Meisam Razaviyayn).

Assumptions. We consider the MSM problem in (4.2) under the following assumptions:

A1) Each set $\mathcal{X}_i$ for $i = 1, \ldots, I$ is (nonempty) closed and convex.

A2) Each set $\Lambda_i \subset \mathbb{R}^{m_i}$ for $i = 1, \ldots, I$ is (nonempty) compact and convex.

A3) Each function $f_{i,j}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, defined on an open convex set Ω containing $\mathcal{X}$, is continuously differentiable on Ω. Moreover, each $f_{i,j}$ is *either* convex or concave on $\mathcal{X}$.

A4) Each function $h_{i,j}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, defined on an open convex set Ω_i containing Λ_i, is continuously differentiable on Ω_i, with the property that the product $h_{i,j}(\boldsymbol{\lambda}_i) f_{i,j}(\mathbf{x})$ is concave in $\boldsymbol{\lambda}_i$ for fixed $\mathbf{x} \in \mathcal{X}$.

A5) Each function s_k for $k = 1, \ldots, n_s$ is continuously differentiable and convex on $\mathcal{X}$.

Clearly, under A1 and A5 the feasible set of the optimization problem (4.2) i.e. $\Xi \triangleq \{\mathbf{x} \in \mathcal{X} : \mathbf{s}(\mathbf{x}) \leq \mathbf{0}\}$ is convex. The MSM problem has some special features that make it challenging, namely:

1. The objective function of the MSM problem is in general nonconcave, and nondifferentiable.

2. The MSM problem can be cast as a Mathematical Program with Equilibrium Constraints (MPEC) [66]. Specifically, it is not difficult to see that the problem in (4.2) is equivalent to:

$$\begin{aligned} \underset{\mathbf{x},\boldsymbol{\lambda}}{\text{maximize}} \quad & \sum_{i=1}^{I} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i) f_{i,j}(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \Xi \\ & \boldsymbol{\lambda}_i \in \underset{\widehat{\boldsymbol{\lambda}}_i \in \Lambda_i}{\text{argmax}} \sum_{j=1}^{J} h_{i,j}(\widehat{\boldsymbol{\lambda}}_i) f_{i,j}(\mathbf{x}) \quad \forall\, i = 1, \ldots, I. \end{aligned} \tag{4.3}$$

Notice that, under the assumptions above, given any $\mathbf{x} \in \Xi$ the optimization problems

$$\underset{\widehat{\boldsymbol{\lambda}}_i \in \Lambda_i}{\text{maximize}} \sum_{j=1}^{J} h_{i,j}(\widehat{\boldsymbol{\lambda}}_i) f_{i,j}(\mathbf{x})$$

are concave for every $i = 1, \ldots, I$. Thus, the last set of constraints in (4.3) is equivalent to each $\boldsymbol{\lambda}_i$, $i = 1, \ldots, I$, belonging to the solution set of a (parametrized) Variational Inequality (VI) [28]; giving rise to a MPEC. In general, the MPECs are considered NP-hard (see e.g., [8]). This fact is noted only to highlight the challenge of the problem (4.3) and it is not used in the subsequent discussion.

Due to the characteristics of the MSM problem mentioned above, it is clear that finding a (globally) optimal solution of (4.2) is intrinsically hard. We then devote the first part of this chapter to study such a class of nonconcave and nondifferentiable maximization problem focusing on finding *solutions of practical interest*, that is, points satisfying the first order optimality conditions of "simpler" optimization problems but still associated with the MSM. It is important to mention that, diverse resource allocation problems in the emerging field of *physical layer based security* (see, e.g., [47] for a recent survey on this topic) serve as the main motivation for us to study the MSM problem. Hence, the main objectives of this chapter are twofold. First, derive easily implementable iterative algorithms (with provable convergence) that attempt to find a solution of (4.2); and second, apply this general problem formulation to signal processing applications.

In pursuance of the first objective stated above, we propose two different approaches. The first one is based on a nontrivial Difference of Concave (DC)[2] representation of the MSM problem; thus, we rely on the widely studied area of DC-Programming (see, e.g., [43, 5] and the references therein). The second approach capitalizes on a (smooth) reformulation of the original problem built upon the joint optimization of the variables present in such a program. Then, we study the relation between the MSM problem and such reformulations. Namely, we carefully define the concepts of stationary points

[2]In this chapter and unless stated the contrary, since we deal with maximization problems, DC stands for Difference of Concave regardless of the fact that a difference of concave functions is also a difference of convex functions.

of the original problem and those for the associated programs, and establish connections between them. Finally, we capitalize on the aforementioned MSM problem's reformulations to derive iterative algorithms with provable convergence to points satisfying the first order optimality conditions of those programs associated with the MSM. More precisely, the DC reformulation of the MSM problem leads naturally to adopt the well-known DCA (Difference of Convex Algorithm), see, e.g., [108, 109, 5, 122, 53, 103]; while, for the smooth reformulation, it is possible to devise iterative algorithms based on Successive Convex Approximation (SCA) techniques (see, e.g., [72, 90, 94, 4]). We remark that the work presented in this chapter contains a rigorous and novel treatment of resource allocation problems in multiuser systems where the utility function is a continuous max function of the form (4.1).

In the second part of this chapter, we apply the aforementioned methods to dynamic spectrum management problems in the context of physical layer based security. Different from current cryptographic techniques used to guarantee security among the network users, the main objective of physical layer security is to exploit the *physical characteristics* of the communication channel in pursuance of secure transmissions. This idea was introduced by Aaron Wyner in [116] where he showed that there exists a non-zero transmission rate, the so-called *secrecy rate*, at which the source and destination can exchange perfectly secure messages, whenever the eavesdropper's channel is a degraded version of the transmitter-receiver's channel. In this chapter, similar to the application considered in Chapter 3, we are interested in the Cooperative Jamming (CJ) paradigm (see, e.g., [24, 39, 59]), where additional entities, named friendly jammers, are introduced into the system with the objective of reducing the eavesdroppers' ability to decode the confidential information transmitted between intended nodes by creating judicious interference.

Different approaches have been proposed in the literature to allocate resources in the aforementioned context; for example, in [115, 104, 36, 119, 4] the interactions between the source and the friendly jammers are investigated within the framework of game theory or auction theory. A common characteristic among the cited references is that the communication between the legitimate parties occurs over a single subchannel, and except from [4], the system models considered are composed of either one source-destination link and (possibly) multiple jammers or multiple source-destination links but one friendly jammer. It is worth mentioning that, in [48, 114], the authors consider the resource allocation problem in the context of physical layer security over *multi-carrier* broadband wireless networks. Different from the models described above, we are interested in the secrecy sum-rate maximization problem for a wireless communication system composed of *multiple* legitimate users, *multiple* friendly jammers and a single eavesdropper, where the legitimate parties communicate over *multiple* subchannels. We study such a problem and its extensions for both the cases of non-orthogonal and orthogonal subchannels. Interestingly, the resulting family of resource allocation problems can be cast into the MSM form in (4.2), and thus we can apply the iterative algorithms developed in the first part of this chapter to attempt their solution. Even though most of our discussion deals with SISO (Single-Input-Single-Output) systems, our results are easily extended to study MIMO (Multiple-Input-Multiple-Output) systems.

The main contributions of this chapter can be summarized as follows. First, we find a nontrivial DC-decomposition for the sum of *continuous* max functions having the structure given in the objective function of (4.2). Such a DC-decomposition leads directly to the development of a class of algorithms with provable convergence for resource allocation problems of the form (4.2). A different approach based on the joint optimization of the variables in the problem (4.2) permits the design of a second class of algorithms for addressing the MSM problem. The third main contribution is a rigorous treatment of the secrecy sum-rate maximization problem in a multi-user, multi-jammer and multi-channel network, and consequent centralized algorithms to attempt its solution. Moreover, for the case of OFDMA (Orthogonal Frequency Division Multiple Access) transmissions, the resulting algorithms are distributed and can be applied to compute relaxed equilibrium points of a game theoretical model [76] used to allocate resources in this setting; thus extending the results in Chapter 3. Finally, iterative algorithms for alternative system designs, such as those involving quality of service constraints or the well-known Max-Min fairness, are also developed in this chapter.

The rest of this chapter is organized as follows. Section 4.2 introduces the DC-Programming approach, while Section 4.3 proposes the joint optimization reformulation technique. Each of these two main sections is divided in two parts: first, Subsections 4.2.1 and 4.3.1 define the stationarity concepts for both the MSM problem and its reformulation; and second, Subsections 4.2.2 and 4.3.2 describe the proposed algorithms and their convergence properties. In order to encompass more applications, Subsection 4.2.3 extends the problem formulation in (4.2) to the case of coupling constraints that are of the DC-type, giving rise to a nonconvex feasible set. The theoretical part of this chapter concludes with Section 4.4, where we contrast both of the approaches proposed to address the MSM program. Section 4.5 presents some applications of the theoretical results to the area of physical layer based security. More precisely, Subsection 4.5.1 deals with the multiple non-orthogonal subchannels scenario, and Subsection 4.5.2 presents a game theoretical model for the case of OFDMA transmissions. Subsection 4.5.3 capitalizes on the theory in Subsection 4.2.3 to derive iterative algorithms for alternative system designs, such as those involving quality of service constraints. Numerical experiments that validate our theoretical findings are presented in Subsection 4.5.4. Finally, Section 4.6 draws the conclusions and discusses future research directions.

4.2 A DC-Programming Approach

With the objective of deriving an iterative algorithm that attempts to find a solution of the MSM problem in (4.2), we introduce the first proposed approach that depends on a nontrivial DC reformulation of such a problem. This gives rise to the first main result of this section, which is stated formally in Proposition 4.1. It is worth mentioning that a well-known result in the DC literature states that the point-wise maximum of a finite set of DC functions is also a function of the DC-type; see, e.g., [42, Thm. 4.1(ii)]. In what follows, we show that θ, a sum of continuous max functions, is of the DC-type. Even though, this result relies critically on the structure of each summand within the maximand in θ_i, our constructive proof below appears to be new in the related literature.

In order to find the aforementioned DC-decomposition of θ, we start by introducing some definitions. For every $i = 1, \ldots, I$ and $j = 1, \ldots, J$ let

$$\rho_{i,j}^{\max} \triangleq \underset{\boldsymbol{\lambda}_i \in \Lambda_i}{\text{maximum}} \; h_{i,j}(\boldsymbol{\lambda}_i) \quad \text{and} \quad \rho_{i,j}^{\min} \triangleq \underset{\boldsymbol{\lambda}_i \in \Lambda_i}{\text{minimum}} \; h_{i,j}(\boldsymbol{\lambda}_i). \tag{4.4}$$

Note that, by Weierstrass' Theorem [11, Prop. 2.1.1], the quantities above are well-defined since each function $h_{i,j}$ is continuous on Λ_i, and each set Λ_i is compact (see assumptions A2 and A4). Using these definitions, we rewrite each product $h_{i,j}(\boldsymbol{\lambda}_i)f_{i,j}(\mathbf{x})$ by distinguishing the following two cases:

(i) if $f_{i,j}(\mathbf{x})$ is convex (cvx) then

$$h_{i,j}(\boldsymbol{\lambda}_i)f_{i,j}(\mathbf{x}) = \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \left(h_{i,j}(\boldsymbol{\lambda}_i) - \rho_{i,j}^{\min}\right) f_{i,j}(\mathbf{x});$$

(ii) if $f_{i,j}(\mathbf{x})$ is concave (cve) then

$$h_{i,j}(\boldsymbol{\lambda}_i)f_{i,j}(\mathbf{x}) = \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) + \left(\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i)\right) (-f_{i,j}(\mathbf{x})).$$

Therefore, each $\theta_i(\mathbf{x})$ [cf. (4.1)] for $i = 1, \ldots, I$ can be rewritten equivalently as

$$\begin{aligned}\theta_i(\mathbf{x}) = &\sum_{j:f_{i,j}\ \text{cvx}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j:f_{i,j}\ \text{cve}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) \\ &+ \underset{\boldsymbol{\lambda}_i \in \Lambda_i}{\text{maximum}} \left(\sum_{j:f_{i,j}\ \text{cvx}} \left(h_{i,j}(\boldsymbol{\lambda}_i) - \rho_{i,j}^{\min}\right) f_{i,j}(\mathbf{x}) + \sum_{j:f_{i,j}\ \text{cve}} \left(\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i)\right) (-f_{i,j}(\mathbf{x})) \right).\end{aligned} \tag{4.5}$$

Remark 4.1. *(On the Properties of the Decomposition of θ_i).* It is important to highlight the following facts with regard to the decomposition of each function $\theta_i(\mathbf{x})$ given in (4.5):

1. If $\rho_{i,j}^{\min} \leq 0$ for some j such that $f_{i,j}$ is convex, then $\rho_{i,j}^{\min} f_{i,j}(\mathbf{x})$ is a *concave* function on $\mathcal{X}$. Of course, if $\rho_{i,j}^{\min} > 0$ such product is a *convex* function on $\mathcal{X}$.

2. If $\rho_{i,j}^{\max} \geq 0$ for some j such that $f_{i,j}$ is concave, then $\rho_{i,j}^{\max} f_{i,j}(\mathbf{x})$ is a *concave* function on $\mathcal{X}$. Of course, if $\rho_{i,j}^{\max} < 0$ such product is a *convex* function on $\mathcal{X}$.

3. For $i = 1, \ldots, I$, let

$$g_i(\mathbf{x}) \triangleq \underset{\boldsymbol{\lambda}_i \in \Lambda_i}{\text{maximum}} \ G_i(\mathbf{x}, \boldsymbol{\lambda}_i), \tag{4.6}$$

where

$$G_i(\mathbf{x}, \boldsymbol{\lambda}_i) \triangleq \sum_{j: f_{i,j} \text{ cvx}} \big(h_{i,j}(\boldsymbol{\lambda}_i) - \rho_{i,j}^{\min}\big) f_{i,j}(\mathbf{x}) + \sum_{j: f_{i,j} \text{ cve}} \big(\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i)\big)(-f_{i,j}(\mathbf{x})). \tag{4.7}$$

Note that, the function G_i has the following two properties: first, $G_i(\bullet, \boldsymbol{\lambda}_i)$ is convex on $\mathcal{X}$ for all $\boldsymbol{\lambda}_i \in \Lambda_i$; and second, $G_i(\mathbf{x}, \bullet)$ is concave on Λ_i for all $\mathbf{x} \in \mathcal{X}$ (refer to assumption A4). As a consequence of the former property of G_i, it is not difficult to show that the function g_i is convex on $\mathcal{X}$ (see, [11, Prop. 4.5.1(a)]).

The observations in Remark 4.1 pave the way to derive a DC-decomposition of each function $\theta_i(\mathbf{x})$. In favor of obtaining such decomposition, let us introduce the following index sets: for every $i = 1, \ldots, I$

$$\begin{aligned} \mathcal{J}_i^{\text{cvx}} &\triangleq \{j : f_{i,j} \text{ is convex and } \rho_{i,j}^{\min} \leq 0\} \\ \overline{\mathcal{J}}_i^{\text{cvx}} &\triangleq \{j : f_{i,j} \text{ is convex and } \rho_{i,j}^{\min} > 0\}; \quad \text{and} \\ \mathcal{J}_i^{\text{cve}} &\triangleq \{j : f_{i,j} \text{ is concave and } \rho_{i,j}^{\max} \geq 0\} \\ \overline{\mathcal{J}}_i^{\text{cve}} &\triangleq \{j : f_{i,j} \text{ is concave and } \rho_{i,j}^{\max} < 0\}. \end{aligned} \tag{4.8}$$

Based on the partition generated by these sets of indices, it is not difficult to observe that (4.5) can be rewritten as follows:

$$\begin{aligned}\theta_i(\mathbf{x}) &= \sum_{j\in\mathcal{J}_i^{\mathrm{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j\in\overline{\mathcal{J}}_i^{\mathrm{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j\in\mathcal{J}_i^{\mathrm{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) \\ &\quad + \sum_{j\in\overline{\mathcal{J}}_i^{\mathrm{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) + g_i(\mathbf{x}) \\ &= \underbrace{\left(\sum_{j\in\mathcal{J}_i^{\mathrm{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j\in\mathcal{J}_i^{\mathrm{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x})\right)}_{\triangleq\, u_i(\mathbf{x})} \\ &\quad - \underbrace{\left(-\sum_{j\in\overline{\mathcal{J}}_i^{\mathrm{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) - \sum_{j\in\overline{\mathcal{J}}_i^{\mathrm{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) - g_i(\mathbf{x})\right)}_{\triangleq\, v_i(\mathbf{x})}\end{aligned} \tag{4.9}$$

where u_i and v_i are clearly concave functions on $\mathcal{X}$. Hence, we have obtained a DC-decomposition of the functions $\theta_i(\mathbf{x})$ for every $i = 1, \ldots, I$. Consequently, since the sum of DC functions is also DC [42, Thm. 4.1(i)], the objective function of the MSM problem in (4.2)

$$\theta(\mathbf{x}) = u(\mathbf{x}) - v(\mathbf{x}), \tag{4.10}$$

where

$$u(\mathbf{x}) \triangleq \sum_{i=1}^{I} u_i(\mathbf{x}) = \sum_{i=1}^{I}\left(\sum_{j\in\mathcal{J}_i^{\mathrm{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j\in\mathcal{J}_i^{\mathrm{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x})\right) \tag{4.11}$$

$$v(\mathbf{x}) \triangleq \sum_{i=1}^{I} v_i(\mathbf{x}) = -\sum_{i=1}^{I}\left(\sum_{j\in\overline{\mathcal{J}}_i^{\mathrm{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j\in\overline{\mathcal{J}}_i^{\mathrm{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) + g_i(\mathbf{x})\right) \tag{4.12}$$

is a function of the DC-type, where the function u is differentiable on $\mathcal{X}$, while v is not. In other words, we have found a nontrivial DC-decomposition of the summation of *continuous max functions* of the form (4.1). This observation yields the main result of this section, which is summarized in the next proposition and whose proof follows readily from the previous construction.

Proposition 4.1. Under assumptions A1-A5, the objective function $\theta(\mathbf{x})$ of the MSM problem in (4.2) is a DC-type function with DC-decomposition $u(\mathbf{x}) - v(\mathbf{x})$, where the concave functions u and v are defined in (4.11) and (4.12), respectively.

A direct consequence of the proposition above is that the MSM problem can be rewritten as the following DC program:

$$(\mathrm{MSM_{DC}}): \quad \underset{\mathbf{x}\in\Xi}{\text{maximize}}\ u(\mathbf{x}) - v(\mathbf{x}). \tag{4.13}$$

This result is key in the derivation of an iterative algorithm for attempting the solution of the nonconvex and nondifferentiable optimization problem (4.2) via its DC-reformulation (4.13). However, it remains to explore in depth the relation between these two problems; we address this issue in the next section and, in Subsection 4.2.2, we construct the aforementioned iterative algorithm.

4.2.1 Stationarity Concepts

In this section, we establish some connections between the MSM problem (4.2) and its DC reformulation (4.13), in terms of (globally and locally) optimal solutions and stationary points. Let us start by introducing the concept of stationary solution for the MSM problem. Note that, a careful definition of this term is required due to the nondifferentiability of the objective function that prevents the use of this classical notion in terms of the gradient; and, the nonconcavity of the objective function restricts also the use of the stationary solution concept in terms of subgradients of concave functions. Under assumptions A1-A5 and by invoking Danskin's Theorem [34, Thm. 1.29], the objective function of the MSM problem is directionally differentiable i.e. in every direction $\mathbf{d} \in \mathbb{R}^n$ the directional derivative of θ is given by

$$\theta'(\mathbf{x}; \mathbf{d}) \triangleq \sum_{i=1}^{I} \left(\underset{\boldsymbol{\lambda}_i \in \Lambda_i^\star(\mathbf{x})}{\text{maximum}} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i) \, \nabla f_{i,j}(\mathbf{x})^T \mathbf{d} \right), \tag{4.14}$$

where, for every $i = 1, \ldots, I$, $\Lambda_i^\star(\mathbf{x})$ denotes the set of maximizing points

$$\Lambda_i^\star(\mathbf{x}) \triangleq \left\{ \boldsymbol{\lambda}_i^\star \in \Lambda_i \ : \ \theta_i(\mathbf{x}) = \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i^\star)\, f_{i,j}(\mathbf{x}) \right\}. \tag{4.15}$$

Notice that, by Weierstrass' theorem [11, Prop. 2.1.1], the sets $\Lambda_i^\star(\mathbf{x}) \neq \emptyset$ for all $i = 1, \ldots, I$. Consequently, we are now able to introduce the concept of stationary solutions for the MSM problem in terms of the directional derivative of θ; for shortness, we call them d-stationary solutions. The following definition formalizes this idea.

Definition 4.1. A point $\mathbf{x}^\star \in \Xi$ is a d-stationary solution of the MSM problem in (4.2) if for all $\mathbf{x} \in \Xi$

$$\theta'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) = \sum_{i=1}^{I} \left(\underset{\boldsymbol{\lambda}_i \in \Lambda_i^\star(\mathbf{x}^\star)}{\text{maximum}} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i)\, \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \right) \leq 0,$$

where the sets $\Lambda_i^\star(\mathbf{x})$ for $i = 1, \ldots, I$ are defined in (4.15).

It is clear that the objective function of the DC problem (4.13) has directional derivatives in every direction $\mathbf{d} \in \mathbb{R}^n$ given by $u'(\mathbf{x}; \mathbf{d}) - v'(\mathbf{x}; \mathbf{d})$ where

$$u'(\mathbf{x}; \mathbf{d}) = \sum_{i=1}^{I} \left(\sum_{j \in \mathcal{J}_i^{\text{cvx}}} \rho_{i,j}^{\min}\, \nabla f_{i,j}(\mathbf{x})^T \mathbf{d} + \sum_{j \in \mathcal{J}_i^{\text{cve}}} \rho_{i,j}^{\max}\, \nabla f_{i,j}(\mathbf{x})^T \mathbf{d} \right) \tag{4.16}$$

and,

$$v'(\mathbf{x}; \mathbf{d}) = -\sum_{i=1}^{I} \left(\sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min}\, \nabla f_{i,j}(\mathbf{x})^T \mathbf{d} + \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max}\, \nabla f_{i,j}(\mathbf{x})^T \mathbf{d} + g_i'(\mathbf{x}; \mathbf{d}) \right). \tag{4.17}$$

Therefore, we can also define a d-stationary solution of the DC problem in (4.13) as follows.

Definition 4.2. A point $\mathbf{x}^\star \in \Xi$ is a d-stationary solution of the DC maximization problem in (4.13) if

$$u'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) - v'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) \leq 0 \quad \forall\, \mathbf{x} \in \Xi.$$

where the directional derivatives u' and v' are defined in (4.16) and (4.17), respectively.

It is worth remarking that, invoking the properties of a DC function (see, e.g., [42, Sec. 4.2] or [40, Eq. 1.1]) we have that the directional derivative of the DC function $\theta(\mathbf{x}) = u(\mathbf{x}) - v(\mathbf{x})$ exists everywhere, i.e. at every $\mathbf{x} \in \Xi$ and in each direction $\mathbf{d} \in \mathbb{R}^n$, it holds

$$\theta'(\mathbf{x}; \mathbf{d}) = u'(\mathbf{x}; \mathbf{d}) - v'(\mathbf{x}; \mathbf{d}). \tag{4.18}$$

A concept widely used in the DC-Programming literature (for unconstrained problems) is that of *critical points* (see, e.g., [40, 5, 74] and the references therein). A vector $\mathbf{x}^\star$ is said to be a *critical point* of the DC function $u(\mathbf{x}) - v(\mathbf{x})$ (or *generalized KKT point* of the problem $\underset{\mathbf{x} \in \mathbb{R}^n}{\text{maximize}}\ u(\mathbf{x}) - v(\mathbf{x})$) if $\partial u(\mathbf{x}^\star) \cap \partial v(\mathbf{x}^\star) \neq \emptyset$, where $\partial u(\mathbf{x}^\star)$ and $\partial v(\mathbf{x}^\star)$ denote the subdifferentials[3] of the functions u and v at $\mathbf{x}^\star$, respectively. This definition is an extension to that one used in the case of differentiable functions u and v and thus, the subdifferentials are a singleton corresponding to the gradients i.e. $\mathbf{x}^\star$ is a critical point of $u(\mathbf{x}) - v(\mathbf{x})$ if $\nabla u(\mathbf{x}^\star) - \nabla v(\mathbf{x}^\star) = \mathbf{0}$. Since we are interested in *constrained* optimization problems, a natural extension of the aforementioned concept for constrained DC programs is given below (see, e.g. [40]).

Definition 4.3. A vector $\mathbf{x}^\star \in \Xi$ is a critical point of the DC maximization problem in (4.13) if

$$\mathbf{0} \in \nabla u(\mathbf{x}^\star) - \partial v(\mathbf{x}^\star) - \mathcal{N}_\Xi(\mathbf{x}^\star),$$

where $\mathcal{N}_\Xi(\mathbf{x}^\star)$ denotes the normal cone to the convex set Ξ at $\mathbf{x}^\star$, that is $\mathcal{N}_\Xi(\mathbf{x}^\star) \triangleq \left\{\mathbf{d} \in \mathbb{R}^n \ : \ \mathbf{d}^T(\mathbf{x} - \mathbf{x}^\star) \leq 0 \ \ \forall\, \mathbf{x} \in \Xi\right\}$.

Before proceeding with our analysis, it is worth stressing some observations with respect to the subdifferentials of the function v; we summarize those observations in the following remark.

[3] Technically speaking, since we are refering to concave functions, the term used should be *superdifferential* instead of *subdifferential*. However, as done in many sources and without loss of precision, we use subdifferential (and subgradient) for both convex and concave functions.

Remark 4.2. *(On the subdifferential $\partial v(\mathbf{x})$).* In order to characterize $\partial v(\mathbf{x})$ for the concave function v [cf. (4.12)], we only need to focus on each term g_i [cf. (4.6)], since the rest of summands in v are differentiable by assumption. Hence, with the objective of finding the subdifferential of each g_i, we observe that: first, from assumption A3 it follows that for every i the functions $G_i(\bullet, \boldsymbol{\lambda}_i)$ are differentiable on $\mathcal{X}$ for each $\boldsymbol{\lambda}_i \in \Lambda_i$; and second, we introduce the following blanket assumption:

A6) For every $i = 1, \ldots, I$, $\nabla_{\mathbf{x}} G_i(\mathbf{x}, \bullet)$ is continuous on Λ_i for each $\mathbf{x} \in \Xi$.

Thus, under assumptions A1-A6, by invoking Danskin's theorem [11, Prop. 4.5.1(b)] we have that: for every $i = 1, \ldots, I$

$$\partial g_i(\mathbf{x}) = \mathbf{conv}\left\{\nabla_{\mathbf{x}} G_i(\mathbf{x}, \boldsymbol{\lambda}_i) \,:\, \boldsymbol{\lambda}_i \in \Lambda_i^\star(\mathbf{x})\right\} \quad \forall\, \mathbf{x} \in \Xi, \tag{4.19}$$

where **conv** denotes the convex hull of the corresponding set, and the sets $\Lambda_i^\star(\mathbf{x})$ are defined in (4.15).

Based on the definitions and observations introduced above, we are now ready to provide an insight into the relation between the MSM problem and its DC-reformulation. Moreover, at this point in the discussion, a natural question to ask is: can we establish any relation between the d-stationary solutions of (4.2) and the critical points of its associated DC program (4.13)? The next proposition gives an answer to this question, and it also summarizes some connections between the problems MSM and MSM_{DC}. See Figure 4.1 for an illustration of those relations.

Proposition 4.2. Under assumptions A1-A6,

(a) The MSM problem in (4.2) is equivalent to the DC maximization problem (4.13) in terms of globally optimal and d-stationary solutions.

(b) If $\mathbf{x}^\star$ is a local maximum of the MSM problem in (4.2), then $\mathbf{x}^\star$ is a d-stationary solution.

(c) If $\mathbf{x}^\star$ is a d-stationary solution of the MSM problem in (4.2), then $\mathbf{x}^\star$ is a critical point of the DC program in (4.13).

(d) If $\mathbf{x}^\star$ is a critical point of the DC program in (4.13) such that, for every $i = 1, \ldots, I$, $\nabla_{\mathbf{x}} G_i(\mathbf{x}^\star, \boldsymbol{\lambda}_i^\star)$ is constant for all $\boldsymbol{\lambda}_i^\star \in \Lambda_i^\star(\mathbf{x}^\star)$, then $\mathbf{x}^\star$ is a d-stationary solution of the MSM problem in (4.2).

Proof. (a) The first assertion follows readily from Proposition 4.1; while, the second claim is a direct consequence of (4.18).

(b) This is clear.

(c) Suppose that $\mathbf{x}^\star \in \Xi$ is a d-stationary solution of the optimization problem (4.2), i.e. for all $\mathbf{x} \in \Xi$

$$\begin{aligned} 0 \geq \theta'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) &\overset{(a)}{=} u'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) - v'(\mathbf{x};^\star \mathbf{x} - \mathbf{x}^\star) \\ &\overset{(b)}{=} \nabla u(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) - \min_{\mu_v(\mathbf{x}^\star) \in \partial v(\mathbf{x}^\star)} \mu_v(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \\ &\overset{(c)}{\geq} \nabla u(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) - \widehat{\mu}_v(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \\ &= \left[\nabla u(\mathbf{x}^\star) - \widehat{\mu}_v(\mathbf{x}^\star)\right]^T (\mathbf{x} - \mathbf{x}^\star), \end{aligned}$$

where in (a) we invoked (4.18); in (b) we applied the differentiability of u, and invoked [91, Thm. 23.4]; while in (c) we took any $\widehat{\mu}_v(\mathbf{x}^\star) \in \partial v(\mathbf{x}^\star)$. As a result,

$$\left[\nabla u(\mathbf{x}^\star) - \widehat{\mu}_v(\mathbf{x}^\star)\right]^T (\mathbf{x} - \mathbf{x}^\star) \leq 0 \quad \forall \mathbf{x} \in \Xi \;\Rightarrow\; \nabla u(\mathbf{x}^\star) - \widehat{\mu}_v(\mathbf{x}^\star) \in \mathcal{N}_\Xi(\mathbf{x}^\star)$$

i.e. $\mathbf{x}^\star$ is a critical point of the DC program in (4.13).

(d) Let $\mathbf{x}^\star \in \Xi$ be a critical point of the DC program (4.13). By assumption, we have that for every $i = 1, \ldots, I$ the gradients $\nabla_{\mathbf{x}} G_i(\mathbf{x}^\star, \boldsymbol{\lambda}_i^\star)$ are constant for all $\boldsymbol{\lambda}_i^\star \in \Lambda_i^\star(\mathbf{x}^\star)$, then from (4.19) it follows that the subdifferential of $\partial g_i(\mathbf{x}^\star) = \{\nabla_{\mathbf{x}} G_i(\mathbf{x}^\star, \boldsymbol{\lambda}_i^\star)\}$, hence, by the differentiability of the rest of summands in v, we have that $\partial v(\mathbf{x}^\star) = \{\nabla v(\mathbf{x}^\star)\}$. Since $\mathbf{x}^\star$ is a critical point of (4.13), then $\mathbf{0} \in \nabla u(\mathbf{x}^\star) - \nabla v(\mathbf{x}^\star) - \mathcal{N}_\Xi(\mathbf{x}^\star)$, that is, for all $\mathbf{x} \in \Xi$

$$\begin{aligned} 0 &\geq \left[\nabla u(\mathbf{x}^\star) - \nabla v(\mathbf{x}^\star)\right]^T (\mathbf{x} - \mathbf{x}^\star) \\ &= u'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) - v'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) \\ &= \theta'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) \end{aligned}$$

where the last equality follows from (4.18). Therefore, $\mathbf{x}^\star$ is a d-stationary solution of (4.2). $\square$

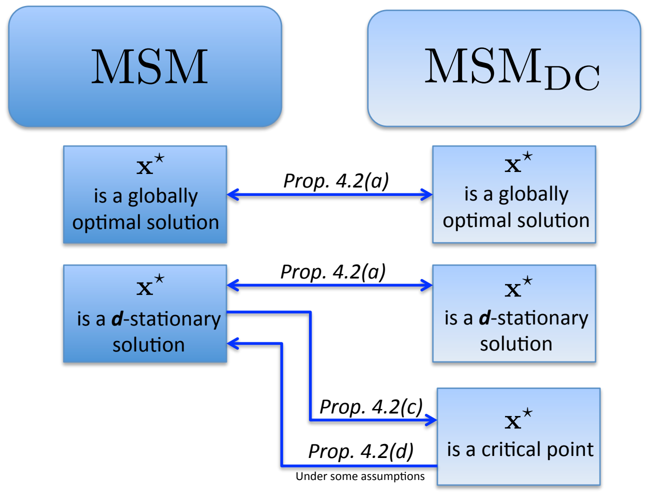

Figure 4.1: Connections between the problems MSM, introduced in (4.2), and MSM$_{\text{DC}}$ defined in (4.13).

Proposition 4.2(c) states that every d-stationary solution of the MSM problem is a critical point of its corresponding DC-reformulation. Nevertheless, the converse is not always true as stated in Proposition 4.2(d). This statement requires that, for every $i = 1, \ldots, I$, the gradients $\nabla_{\mathbf{x}} G_i(\mathbf{x}^\star, \boldsymbol{\lambda}_i^\star)$ are constant for all $\boldsymbol{\lambda}_i^\star \in \Lambda_i^\star(\mathbf{x}^\star)$ so that a critical point $\mathbf{x}^\star \in \Xi$ of the DC problem (4.13) is a d-stationary solution of the MSM problem. Notice that this condition is trivially satisfied when the sets $\Lambda_i^\star(\mathbf{x}^\star)$ are a singleton. Clearly, this condition implies that each g_i, for all $i = 1, \ldots, I$, is differentiable at $\mathbf{x}^\star$; thus, the function v is differentiable at $\mathbf{x}^\star$. A sufficient condition for this to hold is that each $G_i(\mathbf{x}^\star, \bullet)$ is strongly concave on Λ_i for every $i = 1, \ldots, I$. It is worth mentioning, that for those applications that do not satisfy the conditions required by Proposition 4.2(d), it may still be possible to derive some connections between critical points of the DC program and d-stationary solutions of the MSM problem by exploiting the problem's structure; see, for example, the result in Proposition 4.8.

4.2.2 DC-based Algorithm

As already explained, in general, computing a globally optimal solution of the MSM problem (4.2) is challenging. Therefore, in this section, we turn our attention to derive an algorithm that attempts the computation of critical points (in the sense of Definition 4.3) of the MSM$_{\text{DC}}$. The main implication of Proposition 4.1 is that DCA-based schemes, (see, e.g., [108, 109, 5, 122, 53, 103] and the references therein) can be constructed in order to achieve our objective.

The concepts of DC programming and DCA were introduced by Pham (refer to [108, 109, 5]) and have been used extensively in the nonconvex/nonsmooth optimization field. Due to its wide scope, the authors of [56] do not consider the DCA as an algorithm but rather as a philosophy that can be adopted for deriving a range of algorithms applicable in the solution of different problems. In what follows, we adopt this technique towards attempting the solution of the MSM problem.

The main idea of the DCA-based frameworks is to solve a sequence of (strongly) concave problems, wherein the DC-type objective function $\theta(\mathbf{x}) = u(\mathbf{x}) - v(\mathbf{x})$ is replaced by a *concave minorand* constructed at the current iteration $\mathbf{x}^\nu \in \Xi$, which we denote by $\widetilde{\theta}(\mathbf{x}; \mathbf{x}^\nu)$. Essentially, this approximating function $\widetilde{\theta}(\mathbf{x}; \mathbf{x}^\nu)$ is obtained by keeping the concave part unchanged and linearizing at $\mathbf{x}^\nu$ the convex part. More specifically, we construct $\widetilde{\theta}(\mathbf{x}; \mathbf{x}^\nu)$ by following the three steps outlined next:

1. The concave function u [cf. (4.11)] remains unchanged.

2. Each function v_i, for $i = 1, \ldots, I$, is replaced by a linear approximation. In particular, for fixed i and for every $j \in \overline{\mathcal{J}}_i^{\text{cvx}} \cup \overline{\mathcal{J}}_i^{\text{cve}}$ we define the following linear approximation of $f_{i,j}$ at a given $\mathbf{x}^\nu \in \Xi$,

$$f_{i,j}(\mathbf{x}) \approx \widetilde{f}_{i,j}(\mathbf{x}; \mathbf{x}^\nu) \triangleq f_{i,j}(\mathbf{x}^\nu) + \nabla f_{i,j}(\mathbf{x}^\nu)^T(\mathbf{x} - \mathbf{x}^\nu).$$

   Similarly, given $\mathbf{x}^\nu \in \Xi$, let $\mu_{g_i}(\mathbf{x}^\nu) \in \partial g_i(\mathbf{x}^\nu)$ be a subgradient of the convex functions g_i at $\mathbf{x}^\nu$ for all $i = 1, \ldots, I$, then consider the following linear approximation of g_i at $\mathbf{x}^\nu$,

$$g_i(\mathbf{x}) \approx \widetilde{g}_i(\mathbf{x}; \mathbf{x}^\nu) \triangleq g_i(\mathbf{x}^\nu) + \mu_{g_i}(\mathbf{x}^\nu)^T(\mathbf{x} - \mathbf{x}^\nu).$$

   Thus, for every $i = 1, \ldots, I$, each function v_i can be approximated by

$$\widetilde{v}_i(\mathbf{x}; \mathbf{x}^\nu) \triangleq -\sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min} \widetilde{f}_{i,j}(\mathbf{x}; \mathbf{x}^\nu) - \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max} \widetilde{f}_{i,j}(\mathbf{x}; \mathbf{x}^\nu) - \widetilde{g}_i(\mathbf{x}; \mathbf{x}^\nu).$$

3. From the two previous steps, it follows that the candidate approxi-

mation of each function θ_i at a given iteration $\mathbf{x}^\nu \in \Xi$ is the concave function

$$\begin{aligned}\widetilde{\theta}_i(\mathbf{x};\mathbf{x}^\nu) &\triangleq u_i(\mathbf{x}) - \widetilde{v}_i(\mathbf{x};\mathbf{x}^\nu)\\ &= \sum_{j\in\mathcal{J}_i^{\mathrm{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j\in\mathcal{J}_i^{\mathrm{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) + \sum_{j\in\overline{\mathcal{J}}_i^{\mathrm{cvx}}} \rho_{i,j}^{\min} \widetilde{f}_{i,j}(\mathbf{x};\mathbf{x}^\nu) +\\ &\quad + \sum_{j\in\overline{\mathcal{J}}_i^{\mathrm{cve}}} \rho_{i,j}^{\max} \widetilde{f}_{i,j}(\mathbf{x};\mathbf{x}^\nu) + \widetilde{g}_i(\mathbf{x};\mathbf{x}^\nu).\end{aligned} \tag{4.20}$$

Finally, it is not difficult to show that $\theta(\mathbf{x})$ is minorized at $\mathbf{x}^\nu$ over the set Ξ by the following strongly concave function

$$\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu) \triangleq \sum_{i=1}^{I} \widetilde{\theta}_i(\mathbf{x};\mathbf{x}^\nu) - \frac{\tau}{2}\|\mathbf{x}-\mathbf{x}^\nu\|^2. \tag{4.21}$$

where we added a regularization term with $\tau > 0$, whose benefits are well-understood (see, e.g., [12]).

As a result, the proposed DC-based scheme consists in solving iteratively the following sequence of strongly concave problems: given $\mathbf{x}^\nu \in \Xi$

$$\widehat{\mathbf{x}}(\mathbf{x}^\nu) \triangleq \underset{\mathbf{x}\in\Xi}{\operatorname{argmax}}\ \widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu). \tag{4.22}$$

The formal description of the proposed scheme is given in Algorithm 4.1, and its convergence properties are stated in Proposition 4.3.

Algorithm 4.1: DC-based Algorithm for the MSM Problem

Data: $\tau > 0$ and $\mathbf{x}^0 \in \Xi$. Set $\nu = 0$.
(`S.0`): For every $i = 1, \ldots, I$ and $j = 1, \ldots, J$ compute $\rho_{i,j}^{\min}$ and $\rho_{i,j}^{\max}$.
(`S.1`): If $\mathbf{x}^\nu$ satisfies a termination criterion, STOP.
(`S.2`): For $i = 1, \ldots, I$ compute any $\mu_{g_i}(\mathbf{x}^\nu) \in \partial g_i(\mathbf{x}^\nu)$.
(`S.3`): Compute $\widehat{\mathbf{x}}(\mathbf{x}^\nu)$ [cf. (4.22)].
(`S.4`): Set $\mathbf{x}^{\nu+1} \triangleq \widehat{\mathbf{x}}(\mathbf{x}^\nu)$.
(`S.5`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

Remark 4.3. *(On Algorithm 4.1).* A practical termination criterion in Step (`S.1`) of Algorithm 4.1 is to stop the iterates when $|\theta(\mathbf{x}^\nu) - \theta(\mathbf{x}^{\nu-1})| \leq \delta$, where δ is a prescribed accuracy. Notice that, Step (`S.2`) of Algorithm 4.1 requires the computation of $\mu_{g_i}(\mathbf{x}^\nu) \in \partial g_i(\mathbf{x}^\nu)$ for every $i = 1, \ldots, I$, refer to equation (4.19) where these subdifferential sets are characterized. From (4.19), it follows that this step implicitly requires the solution of the following concave maximization problems: given $\mathbf{x}^\nu \in \Xi$, for $i = 1, \ldots, I$

$$\boldsymbol{\lambda}_i^{\star,\nu} \in \underset{\boldsymbol{\lambda}_i \in \Lambda_i}{\operatorname{argmax}} \; G_i(\mathbf{x}^\nu, \boldsymbol{\lambda}_i). \tag{4.23}$$

We emphasize that the difficulty associated with solving these subproblems depends on the particular application of interest. For example, in the cases considered in Section 4.5, the problem in (4.23) reduces to solving a scalar linear maximization program whose closed form solution is given in (4.55). Hence, in these applications, Step (`S.2`) is computationally inexpensive.

The next lemma summarizes some important properties of the map $\Xi \ni \mathbf{y} \mapsto \widehat{\mathbf{x}}(\mathbf{y}) \in \Xi$ defined in (4.22), that are instrumentals to the convergence proof of Algorithm 4.1 as stated in Proposition 4.3 below. In essence, this lemma shows the well-definiteness and continuity of the aforementioned map, as well as a correspondence between the critical points of (4.13) and the fixed points of this single-valued map.

Lemma 4.1. Under assumptions A1-A6, the map $\Xi \ni \mathbf{y} \mapsto \widehat{\mathbf{x}}(\mathbf{y}) \in \Xi$ defined in (4.22) has the following properties:

(a) The map $\widehat{\mathbf{x}}(\bullet)$ is well-defined;

(b) The map $\widehat{\mathbf{x}}(\bullet)$ is continuous on Ξ; and,

(c) The set of fixed points of the map $\widehat{\mathbf{x}}(\bullet)$ coincides with the set of critical points of the DC problem in (4.13).

Proof. (a) Given any $\mathbf{y} \in \Xi$, $\widehat{\mathbf{x}}(\mathbf{y})$ is the unique solution of the strongly concave problem (4.22), as a result the map $\widehat{\mathbf{x}}(\bullet)$ is well-defined.

(b) This is a consequence of Berge's Maximum Theorem [9].

(c) Suppose that $\mathbf{x}^\star$ is a fixed point of the map $\widehat{\mathbf{x}}(\bullet)$ i.e. $\mathbf{x}^\star = \widehat{\mathbf{x}}(\mathbf{x}^\star)$, by definition we have:

$$(\mathbf{x} - \widehat{\mathbf{x}}(\mathbf{x}^\star))^T \, \nabla\widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{x}^\star); \mathbf{x}^\star) \le 0 \quad \forall\, \mathbf{x} \in \Xi, \tag{4.24}$$

where

$$\nabla \widetilde{\theta}(\widehat{\mathbf{x}}(\mathbf{x}^\star); \mathbf{x}^\star) = \sum_{i=1}^{I} \nabla u_i(\widehat{\mathbf{x}}(\mathbf{x}^\star)) - \nabla \widetilde{v}_i(\widehat{\mathbf{x}}(\mathbf{x}^\star); \mathbf{x}^\star) - \tau(\widehat{\mathbf{x}}(\mathbf{x}^\star) - \mathbf{x}^\star),$$

and

$$\begin{aligned}
\nabla u_i(\widehat{\mathbf{x}}(\mathbf{x}^\star)) &= \sum_{j \in \mathcal{J}_i^{\text{cvx}}} \rho_{i,j}^{\min} \nabla f_{i,j}(\widehat{\mathbf{x}}(\mathbf{x}^\star)) + \sum_{j \in \mathcal{J}_i^{\text{cve}}} \rho_{i,j}^{\max} \nabla f_{i,j}(\widehat{\mathbf{x}}(\mathbf{x}^\star)) \\
\nabla \widetilde{v}_i(\widehat{\mathbf{x}}(\mathbf{x}^\star); \mathbf{x}^\star) &= - \sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min} \nabla f_{i,j}(\mathbf{x}^\star) - \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max} \nabla f_{i,j}(\mathbf{x}^\star) - \mu_{g_i}(\mathbf{x}^\star).
\end{aligned}$$

By using the fact that $\widehat{\mathbf{x}}(\mathbf{x}^\star) = \mathbf{x}^\star$ in (4.24), we get: $\forall\, \mathbf{x} \in \Xi$

$$\begin{aligned}
0 &\geq (\mathbf{x} - \mathbf{x}^\star)^T \left[\sum_{i=1}^{I} \nabla u_i(\mathbf{x}^\star) - \nabla \widetilde{v}_i(\mathbf{x}^\star; \mathbf{x}^\star) \right] \\
&= (\mathbf{x} - \mathbf{x}^\star)^T \left[\sum_{i=1}^{I} \nabla u_i(\mathbf{x}^\star) - \left(- \sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min} \nabla f_{i,j}(\mathbf{x}^\star) - \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max} \nabla f_{i,j}(\mathbf{x}^\star) - \mu_{g_i}(\mathbf{x}^\star) \right) \right]
\end{aligned}$$

Invoking [91, Thm. 23.8] and letting, for every $i = 1, \ldots, I$,

$$\mu_{v_i}(\mathbf{x}^\star) \triangleq - \sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min} \nabla f_{i,j}(\mathbf{x}^\star) - \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max} \nabla f_{i,j}(\mathbf{x}^\star) - \mu_{g_i}(\mathbf{x}^\star),$$

it is easy to see that $\mu_{v_i}(\mathbf{x}^\star) \in \partial v_i(\mathbf{x}^\star)$. Thus, for all $\mathbf{x} \in \Xi$

$$\begin{aligned}
0 &\geq (\mathbf{x} - \mathbf{x}^\star)^T \, \left[\sum_{i=1}^{I} \nabla u_i(\mathbf{x}^\star) - \mu_{v_i}(\mathbf{x}^\star) \right] \\
&= (\mathbf{x} - \mathbf{x}^\star)^T \, \left[\nabla u(\mathbf{x}^\star) - \mu_v(\mathbf{x}^\star) \right]. \qquad (4.25)
\end{aligned}$$

where $\nabla u(\mathbf{x}^\star) \triangleq \sum_{i=1}^{I} \nabla u_i(\mathbf{x}^\star)$ and $\mu_v(\mathbf{x}^\star) \triangleq \sum_{i=1}^{I} \mu_{v_i}(\mathbf{x}^\star)$. Following the same argument as above, it is clear that $\mu_v(\mathbf{x}^\star) \in \partial v(\mathbf{x}^\star)$. As a result, from (4.25) it follows that $\nabla u(\mathbf{x}^\star) - \mu_v(\mathbf{x}^\star) \in \mathcal{N}_\Xi(\mathbf{x}^\star)$ i.e. $\mathbf{x}^\star$ is a critical point of (4.13).

For the converse argument, suppose that $\mathbf{x}^\star$ is a critical point of (4.13). It suffices to observe that: first, $\widehat{\mathbf{x}}(\mathbf{x}^\star)$ is the unique solution of (4.22) where we take $\mathbf{y} = \mathbf{x}^\star$; and, second $\mathbf{x}^\star$ is also an optimal solution of (4.22) since it satisfies the variational principle. Therefore, $\mathbf{x}^\star = \widehat{\mathbf{x}}(\mathbf{x}^\star)$ i.e. $\mathbf{x}^\star$ is a fixed point of the map $\widehat{\mathbf{x}}(\bullet)$. $\square$

The next proposition summarizes the convergence of Algorithm 4.1 to a critical point of the DC-representation of the MSM problem given in (4.13). We refer the reader to [108, 109, 5] for the convergence properties and corresponding proof of the generic DCA. Under some conditions, the DCA is shown to be (globally) convergent to a critical point of an unconstrained DC problem. Basically, this convergence proof is based on DC duality theory. The convergence proof of the Concave-Convex Procedure (CCCP) [122], a variant of the DCA applied to the case where v is assumed to be differentiable (and thus not applicable to our case), can be found in [53, 103]. This proof is based on the Zangwill's global convergence theory [123]. In the proof of the following proposition, we deviate from these approaches and rather exploit the particular structure and characteristics of the MSM problem under consideration. It is important to emphasize that our convergence results are in accordance with those in the related literature, see e.g., [89]. As a side note, we refer the interested reader to [106, 75, 70] where other approaches, different from the DCA and CCCP, are proposed to address unconstrained DC programs, such as proximal point like algorithms.

Proposition 4.3. Under assumptions A1-A6, for every initial point $\mathbf{x}^0 \in \Xi$, the sequence $\{\mathbf{x}^\nu\}$ produced by Algorithm 4.1 is well-defined. Moreover, if θ is bounded above on Ξ and for any $\tau > 0$, then every accumulation point of $\{\mathbf{x}^\nu\}$, if it exists, is a critical point of the DC problem (4.13).

Proof. The first assertion of the proposition is clear (see, Lemma 4.1(a)).

For the second part, by induction, since $\mathbf{x}^\nu \in \Xi$, we have

$$\begin{aligned}
\theta(\mathbf{x}^\nu) = \widetilde{\theta}(\mathbf{x}^\nu; \mathbf{x}^\nu) &\overset{(a)}{\leq} \widetilde{\theta}(\mathbf{x}^{\nu+1}; \mathbf{x}^\nu) \\
&= \sum_{i=1}^{I} u_i(\mathbf{x}^{\nu+1}) - \widetilde{v}_i(\mathbf{x}^{\nu+1}; \mathbf{x}^\nu) - \frac{\tau}{2}\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 \\
&\overset{(b)}{\leq} \sum_{i=1}^{I} u_i(\mathbf{x}^{\nu+1}) - v_i(\mathbf{x}^{\nu+1}) - \frac{\tau}{2}\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 \\
&= \theta(\mathbf{x}^{\nu+1}) - \frac{\tau}{2}\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2
\end{aligned}$$

where (a) follows from the definition of $\mathbf{x}^{\nu+1}$ and (b) by using the fact that each $-\widetilde{v}_i$ is an affine minorand of each function $-v_i$ at $\mathbf{x}^\nu$ over Ξ. Therefore, under the assumption that θ is bounded above on Ξ and $\tau > 0$, the sequence $\{\theta(\mathbf{x}^\nu)\}$ converges and

$$\lim_{\nu\to\infty} \|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 = 0. \tag{4.26}$$

Finally, if $\mathbf{x}^\infty \in \Xi$ is an accumulation point of the sequence $\{\mathbf{x}^\nu\}$. Invoking the continuity of $\widehat{\mathbf{x}}(\bullet)$ [see, Lemma 4.1(b)], and from (4.26) using the fact that $\mathbf{x}^{\nu+1} = \widehat{\mathbf{x}}(\mathbf{x}^\nu)$ i.e. $\lim\limits_{\nu\to\infty} \|\widehat{\mathbf{x}}(\mathbf{x}^\nu) - \mathbf{x}^\nu\| = 0$, we must have that $\widehat{\mathbf{x}}(\mathbf{x}^\infty) = \mathbf{x}^\infty$. By Lemma 4.1(c), we have that the set of fixed points of $\widehat{\mathbf{x}}(\bullet)$ coincides with the set of critical points of (4.13); hence $\mathbf{x}^\infty$ is a critical point of the DC problem (4.13). □

To conclude this section, we highlight the importance of the DC reformulation of the MSM problem. Namely, the benefits of such a reformulation are twofold: first, it simplifies the analysis of the MSM problem because it permits the application of the concept of critical points (in Definition 4.3), which is not directly applicable to the original formulation; and second, it leads directly to the construction of Algorithm 4.1, a DC-based iterative scheme, with provable convergence to those critical points. Furthermore, under the conditions of Proposition 4.2(d), the critical points produced by Algorithm 4.1 coincide with the d-stationary solutions of the original MSM problem.

4.2.3 An Extension to the MSM_{DC} Problem

So far, we have considered the MSM problem in (4.2) under the assumption that the set of constraints is convex. Nevertheless, in some applications this assumption is not fulfilled; refer to Subsection 4.5.3 for some examples. As a result, in this subsection we briefly consider a version of the MSM problem under a more general setting. Namely, we drop the convexity assumption of the coupling constraints $\mathbf{s}(\mathbf{x}) \leq \mathbf{0}$, that is, we replace assumption A5 with:

A5′) Each function s_k is of the DC-type, with DC-decomposition given by: $s_k(\mathbf{x}) \triangleq \widehat{u}_k(\mathbf{x}) - \widehat{v}_k(\mathbf{x})$ for every $k = 1, \ldots, n_s$.

In Section 4.2, we found a DC-decomposition for the objective function of the MSM problem, i.e. $\theta(\mathbf{x}) = u(\mathbf{x}) - v(\mathbf{x})$, where the functions u and v

are defined in (4.11) and (4.12), respectively. This result gave rise to the problem MSM_{DC} defined in (4.13). Hence, throughout this section, we turn our attention to the MSM_{DC} problem under A5$'$, i.e.

$$\begin{array}{lll} \underset{\mathbf{x}\triangleq(\mathbf{x}_i)_{i=1}^{I}}{\text{maximize}} & u(\mathbf{x})-v(\mathbf{x}) & \\ \text{subject to} & \mathbf{x}_i\in\mathcal{X}_i \quad \forall\, i=1,\ldots,I & \textit{(convex private constraints)} \\ & \left(\widehat{u}_k(\mathbf{x})-\widehat{v}_k(\mathbf{x})\right)_{k=1}^{n_s}\leq\mathbf{0} & \textit{(DC-type coupling constraints).} \end{array} \tag{4.27}$$

Clearly, we are in presence of a nonconcave and nondifferentiable optimization problem where both the objective function and the coupling constraints are of the DC-type, while the convex constraints are retained in the set $\mathcal{X}$. This category of problems are referred in the DC literature as *general DC programs*. For the sake of notational simplicity, let

$$\Xi^{\mathsf{DC}}\triangleq\{\mathbf{x}\in\mathcal{X}\,:\,\widehat{u}_k(\mathbf{x})-\widehat{v}_k(\mathbf{x})\leq 0 \quad k=1,\ldots,n_s\}$$

denote the nonconvex feasible set of (4.27).

The main objective of this section is to devise an iterative algorithm converging to critical points of (4.27). The concept of critical point for the maximization problem (4.27) is introduced formally in the next definition, where the complementarity notation $0\leq a\perp b\geq 0$ means $a\cdot b=0$, $a\geq 0$ and $b\geq 0$. This concept is taken directly from the literature for general DC programs; see, e.g., [32, 89].

Definition 4.4. If the pair $\left(\mathbf{x}^\star,\boldsymbol{\pi}^\star\triangleq(\pi_k^\star)_{k=1}^{n_s}\right)$ is such that:

$$\begin{aligned} \mathbf{0}\,\in\,\nabla u(\mathbf{x}^\star)-\partial v(\mathbf{x}^\star)-\sum_{k=1}^{n_s}\pi_k^\star\left[\partial\widehat{u}_k(\mathbf{x}^\star)-\partial\widehat{v}_k(\mathbf{x}^\star)\right]-\mathcal{N}_{\mathcal{X}}(\mathbf{x}^\star),& \\ 0\leq\pi_k^\star \quad\perp\quad \widehat{v}_k(\mathbf{x}^\star)-\widehat{u}_k(\mathbf{x}^\star)\geq 0 \quad k=1,\ldots,n_s,& \end{aligned}$$

where $\mathcal{N}_{\mathcal{X}}(\mathbf{x}^\star)$ is the normal cone to the set $\mathcal{X}$ at $\mathbf{x}^\star$. Then, $\mathbf{x}^\star$ is a critical point of the DC problem (4.27) and $\boldsymbol{\pi}^\star$ is the corresponding vector of multipliers for the DC-type coupling constraints.

General DC programs have been studied extensively in the literature. Basically, two approaches have been proposed to attempt the solution of these sort of problems. First, in [108, 55, 5, 57] (exact) penalty functions are used to deal with the DC constraints, and then, the DCA is applied to the resulting (unconstrained) penalized problem. However, some drawbacks of this approach are that the penalty parameter is in general unknown [56] and, in terms of performance, it may lead to a slow convergence of the resulting iterative scheme [89]. A different approach is based on SCA techniques; see, e.g., [122, 101, 53, 89] for some results dealing with the application of this idea to general DC programs. The objective of SCA techniques is to solve a sequence of concave problems obtained by linearizing (at the current iteration) the convex part of the DC function. In the case of a DC constrained problem, this affine approximation is also used to "convexify" such constraints. Toward finding critical points of (4.27) in the sense of Definition 4.4, we follow the latter approach. In particular, the sequence of concave problems is obtained by: given $\mathbf{x}^\nu \in \Xi^{\mathsf{DC}}$

1. The objective function of (4.27) is approximated by the strongly concave function $\widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu)$ defined in (4.21).

2. Each DC-type constraint s_k is replaced by a linear approximation. Let $\mu_{\widehat{u}_k}(\mathbf{x}^\nu) \in \partial\widehat{u}_k(\mathbf{x}^\nu)$ be a subgradient of the concave function $\widehat{u}_k$ at $\mathbf{x}^\nu$, then for $k = 1\ldots,n_s$

$$s_k(\mathbf{x}) \approx \widetilde{s}_k(\mathbf{x};\mathbf{x}^\nu) \triangleq \widehat{u}_k(\mathbf{x}^\nu) + \mu_{\widehat{u}_k}(\mathbf{x}^\nu)^T(\mathbf{x}-\mathbf{x}^\nu) - \widehat{v}_k(\mathbf{x}).$$

As a result, the SCA-based algorithm attempting to compute critical points of (4.27) consists in solving iteratively the following sequence of strongly concave optimization problems: given $\mathbf{x}^\nu \in \Xi^{\mathsf{DC}}$

$$\widehat{\mathbf{x}}^{\mathsf{DC}}(\mathbf{x}^\nu) \triangleq \underset{\mathbf{x}\,\in\,\widetilde{\Xi}^{\mathsf{DC}}(\mathbf{x}^\nu)}{\operatorname{argmax}} \;\; \widetilde{\theta}(\mathbf{x};\mathbf{x}^\nu), \tag{4.28}$$

where $\widetilde{\Xi}^{\mathsf{DC}}(\mathbf{x}^\nu) \triangleq \{\mathbf{x}\in\mathcal{X} \,:\, \widetilde{s}_k(\mathbf{x};\mathbf{x}^\nu) \le 0 \quad \forall\, k = 1,\ldots,n_s\}$.

The description of the proposed scheme is given in Algorithm 4.2, and its convergence properties are stated formally in Proposition 4.4.

Proposition 4.4. Under assumptions A1-A6 with A5 replaced by A5$'$, for every initial point $\mathbf{x}^0 \in \Xi^{\mathsf{DC}}$ the following statements hold:

(a) *Iterates feasibility* – Under a suitable constraint qualification, the sequence $\{(\mathbf{x}^\nu, \boldsymbol{\pi}^\nu)\}$ produced by Algorithm 4.2, is well-defined. Moreover, $\widetilde{\Xi}^{\mathtt{DC}}(\mathbf{x}^\nu) \subseteq \Xi^{\mathtt{DC}}$ i.e. the sequence $\{\mathbf{x}^\nu\}$ is feasible to the MSM_{DC} problem with DC constraints (4.27).

(b) *Convergence* – If θ is bounded above on $\Xi^{\mathtt{DC}}$ and for any $\tau > 0$ then, under a suitable constraint qualification, every accumulation point of the sequence $\{(\mathbf{x}^\nu, \boldsymbol{\pi}^\nu)\}$, if it exists, corresponds to a critical point of the MSM_{DC} problem in (4.27) and to a multiplier of the DC constraints, respectively.

Proof. (a) The well-definiteness of the sequence $\{(\mathbf{x}^\nu, \boldsymbol{\pi}^\nu)\}$ is clear. For the second assertion, given any $\mathbf{x}^\nu \in \Xi^{\mathtt{DC}}$ and for every $k = 1, \ldots, n_s$, since $\mu_{\widehat{u}_k}(\mathbf{x}^\nu) \in \partial \widehat{u}_k(\mathbf{x}^\nu)$ then $\widehat{u}_k(\mathbf{x}^\nu) + \mu_{\widehat{u}_k}(\mathbf{x}^\nu)^T(\mathbf{x} - \mathbf{x}^\nu) \geq \widehat{u}_k(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$. This implies that $s_k(\mathbf{x}) \leq \widetilde{s}_k(\mathbf{x}; \mathbf{x}^\nu)$ for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{x}^\nu \in \Xi^{\mathtt{DC}}$, where the desired result follows immediately.

(b) Follows readily by applying [89, Thm. 2] to our problem formulation in (4.27). $\square$

It is worth stressing that the regularity condition required by Proposition 4.4 is to guarantee the existence of Lagrange multipliers associated with the (convex) coupling constraints of the optimization problem (4.28). An example of a regularity condition is the well-known Slater's constraint qualification (CQ), we refer the reader to [28, Sec. 3.2] for a detailed discussion on this topic.

Algorithm 4.2: SCA-based Algorithm for the MSM_{DC} Problem with DC-type Coupling Constraints (4.27)

Data: $\tau > 0$ and $\mathbf{x}^0 \in \Xi^{\mathtt{DC}}$. Set $\nu = 0$.

(`S.0`): For every $i = 1, \ldots, I$ and $j = 1, \ldots, J$ compute $\rho_{i,j}^{\min}$ and $\rho_{i,j}^{\max}$.

(`S.1`): If $\mathbf{x}^\nu$ satisfies a termination criterion, STOP.

(`S.2`): For $i = 1, \ldots, I$ compute $\mu_{g_i}(\mathbf{x}^\nu) \in \partial g_i(\mathbf{x}^\nu)$, and for $k = 1, \ldots, n_s$ compute $\mu_{\widehat{v}_k}(\mathbf{x}^\nu) \in \partial \widehat{v}_k(\mathbf{x}^\nu)$.

(`S.3`): Compute $\widehat{\mathbf{x}}^{\mathtt{DC}}(\mathbf{x}^\nu)$ [cf. (4.28)] and the corresponding multiplier $\boldsymbol{\pi}^\nu$ of the DC-constraints.

(`S.4`): Set $\mathbf{x}^{\nu+1} \triangleq \widehat{\mathbf{x}}^{\mathtt{DC}}(\mathbf{x}^\nu)$.

(`S.5`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

4.3 A Joint Optimization Approach

In this section, we deviate from the DC-Programming approach, developed throughout Section 4.2, and we propose an alternative set of techniques to attempt the solution of the MSM problem (4.2). The main idea of the approach proposed here lies in the simple observation of optimizing *jointly* the variables of the MSM problem $\mathbf{x}$ and $\boldsymbol{\lambda} \triangleq (\boldsymbol{\lambda}_i)_{i=1}^I$. More precisely, we reformulate the MSM problem as the following joint optimization (JO) program i.e.

$$(\mathrm{MSM_{JO}}) : \quad \begin{array}{ll} \underset{\mathbf{x},\boldsymbol{\lambda}}{\text{maximize}} & \theta^{\mathrm{J}}(\mathbf{x},\boldsymbol{\lambda}) \triangleq \displaystyle\sum_{i=1}^{I}\sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i) f_{i,j}(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \Xi \\ & \boldsymbol{\lambda} \in \Lambda \triangleq \prod_{i=1}^{I} \Lambda_i. \end{array} \tag{4.29}$$

For the sake of notational simplicity, let $\Xi^J \triangleq \{(\mathbf{x},\boldsymbol{\lambda}) : \mathbf{x} \in \Xi, \boldsymbol{\lambda} \in \Lambda\}$ denote the convex feasible set of (4.29). It is worth emphasizing that, in this section, we are attaching the superscript J in some of the notation to distinguish it from the definitions introduced in the previous section.

A great advantage of the joint optimization reformulation (4.29) is that it removes the nondifferentiability issue of the objective function θ of the MSM problem, however θ^{J} is still a nonconcave function.

4.3.1 Stationarity Concepts

At this point in the discussion, it is worth asking: what is the relation between the MSM problem and its joint optimization reformulation $\mathrm{MSM_{JO}}$ in (4.29)? Proposition 4.5 answers this question by summarizing some connections between these problems in terms of globally optimal solutions and stationary points. For the sake of clarity, Figure 4.2 illustrates these results. Since the objective function of the problem (4.29) is differentiable, the usual definition of stationary solution in terms of the gradient applies, thus we omit it. We recall that the concept of d-stationary solution for the MSM problem was introduced in Definition 4.1. As a side note, in the following discussion, we let $\boldsymbol{\lambda}^\star \triangleq (\boldsymbol{\lambda}_i^\star)_{i=1}^I$ and $\Lambda^\star(\mathbf{x}^\star) \triangleq \prod_{i=1}^I \Lambda_i^\star(\mathbf{x}^\star)$.

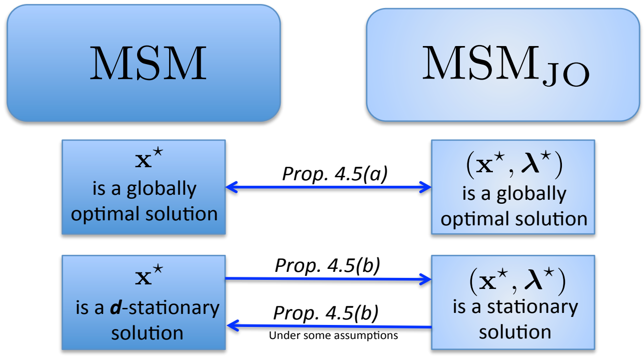

Figure 4.2: Connections between the problems MSM, introduced in (4.2), and $\mathrm{MSM_{DC}}$ defined in (4.29).

Proposition 4.5. Under assumptions A1-A5,

(a) The vector $\mathbf{x}^\star$ is a globally optimal solution of the MSM problem in (4.2) if and only if there exists $\boldsymbol{\lambda}^\star \in \Lambda$ such that the pair $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ is a globally optimal solution of the joint optimization problem in (4.29).

(b) The vector $\mathbf{x}^\star$ is a d-stationary solution of the MSM problem in (4.2) if and only if, for all $\boldsymbol{\lambda}^\star \in \Lambda^\star(\mathbf{x}^\star)$, the pair $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ is a stationary solution of the joint optimization problem in (4.29).

Proof. (a) This assertion is clear.

(b) First, suppose that $\mathbf{x}^\star \in \Xi$ is a d-stationary solution of (4.2), i.e. for all $\mathbf{x} \in \Xi$

$$\sum_{i=1}^{I} \left(\underset{\boldsymbol{\lambda}_i \in \Lambda_i^\star(\mathbf{x}^\star)}{\text{maximum}} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i)\, \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \right) \leq 0$$

First, from assumptions A2 and A4, it is clear that, given $\mathbf{x}^\star \in \Xi$, the sets $\Lambda_i^\star(\mathbf{x}^\star)$ [cf. (4.15)] are non-empty for every $i = 1, \ldots, I$. Therefore, for every $i = 1, \ldots, I$ and for all $\boldsymbol{\lambda}_i^\star \in \Lambda_i^\star(\mathbf{x}^\star)$, by definition it must hold that

$$\sum_{j=1}^{J} f_{i,j}(\mathbf{x}^\star)\, \nabla h_{i,j}(\boldsymbol{\lambda}_i^\star)^T (\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_i^\star) \leq 0 \qquad \forall\, \boldsymbol{\lambda}_i \in \Lambda_i. \tag{4.30}$$

Second, it is easy to see that: for all $\mathbf{x} \in \Xi$ and for all $\boldsymbol{\lambda}^\star \in \Lambda^\star(\mathbf{x}^\star)$

$$
\begin{aligned}
0 &\geq \sum_{i=1}^{I} \left(\underset{\boldsymbol{\lambda}_i \in \Lambda_i^\star(\mathbf{x}^\star)}{\text{maximum}} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i)\, \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \right) \\
&\geq \sum_{i=1}^{I} \left(\sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i^\star)\, \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \right).
\end{aligned} \tag{4.31}
$$

Finally, combining (4.30) and (4.31), it is clear that: for all $\mathbf{x} \in \Xi$, all $\boldsymbol{\lambda} \in \Lambda$ and all $\boldsymbol{\lambda}^\star \in \Lambda^\star(\mathbf{x}^\star)$ it holds that

$$
\begin{aligned}
0 &\geq \sum_{i=1}^{I} \left(\sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i^\star) \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) + \sum_{j=1}^{J} f_{i,j}(\mathbf{x}^\star) \nabla h_{i,j}(\boldsymbol{\lambda}_i^\star)^T (\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_i^\star) \right) \\
&= (\mathbf{x} - \mathbf{x}^\star)^T \nabla_{\mathbf{x}} \theta^J(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) + (\boldsymbol{\lambda} - \boldsymbol{\lambda}^\star)^T \nabla_{\boldsymbol{\lambda}} \theta^J(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) \\
&= [(\mathbf{x}, \boldsymbol{\lambda}) - (\mathbf{x}^\star, \boldsymbol{\lambda}^\star)]^T \nabla \theta^J(\mathbf{x}^\star, \boldsymbol{\lambda}^\star),
\end{aligned} \tag{4.32}
$$

i.e. the pair $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ is a stationary solution of (4.29) for all $\boldsymbol{\lambda}^\star \in \Lambda^\star(\mathbf{x}^\star)$.

For the converse argument, suppose that, for all $\boldsymbol{\lambda}^\star \in \Lambda^\star(\mathbf{x}^\star)$, the pair $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) \in \Xi^J$ is a stationary solution of (4.29) i.e. the inequality in (4.32) holds for any pair $(\mathbf{x}, \boldsymbol{\lambda}) \in \Xi^J$. Thus, if we let $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\star$ in (4.32) it follows that: for all $\mathbf{x} \in \Xi$ and all $\boldsymbol{\lambda}^\star \in \Lambda^\star(\mathbf{x}^\star)$

$$
\sum_{i=1}^{I} \left(\sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i^\star)\, \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \right) \leq 0.
$$

Hence, since the inequality above holds for all $\boldsymbol{\lambda}_i^\star \in \Lambda_i^\star(\mathbf{x}^\star)$ and all $i = 1, \ldots, I$, it is not difficult to see that

$$
\theta'(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) = \sum_{i=1}^{I} \left(\underset{\boldsymbol{\lambda}_i \in \Lambda_i^\star(\mathbf{x}^\star)}{\text{maximum}} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i)\, \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \right) \leq 0
$$

i.e. $\mathbf{x}^\star$ is a d-stationary solution of (4.2). $\square$

It is important to emphasize some aspects with regard to Proposition 4.5(b). First, it follows that if $\mathbf{x}^\star$ is a d-stationary solution of the MSM problem in (4.2), then there must exist $\boldsymbol{\lambda}^\star \in \Lambda$ such that the pair $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ is a stationary solution of the joint optimization problem in (4.29). However, notice that the converse is more stringent, the pair $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ must be a stationary solution of (4.29) for all $\boldsymbol{\lambda}^\star \in \Lambda^\star(\mathbf{x}^\star)$ in order to coincide with a d-stationary solution solution of the MSM problem. The aforementioned condition is readily satisfied when the sets $\Lambda_i^\star(\mathbf{x}^\star)$ are a singleton for all $i = 1, \ldots, I$.

4.3.2 SCA-based Algorithm

The smooth joint optimization reformulation of the MSM problem in (4.29) leads directly to the design of an iterative algorithm attempting the solution of the MSM problem. More precisely, we extend the results in Chapter 3 toward the construction of such schemes converging to stationary solutions of the MSM_{JO}.

Using the definitions of the scalars $\rho_{i,j}^{\max}$ and $\rho_{i,j}^{\min}$ given in (4.4), and those of the index sets $\mathcal{J}_i^{\text{cvx}}$, $\mathcal{J}_i^{\text{cve}}$ and their respective complements in (4.8), it is not difficult to observe that

$$\begin{aligned}
\theta_i^{\text{J}}(\mathbf{x}, \boldsymbol{\lambda}_i) &\triangleq \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i) f_{i,j}(\mathbf{x}) \\
&= \underbrace{\left(\sum_{j \in \mathcal{J}_i^{\text{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j \in \mathcal{J}_i^{\text{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) \right)}_{\text{concave on } \mathcal{X}} \\
&\quad + \underbrace{\left(\sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}) \right)}_{\text{convex on } \mathcal{X}} + \underbrace{G_i(\mathbf{x}, \boldsymbol{\lambda}_i),}_{\text{concave on } \Lambda_i, \text{ convex on } \mathcal{X}}
\end{aligned} \tag{4.33}$$

where G_i is defined in (4.7). Here we propose a SCA-based algorithm. This scheme consists in solving a sequence of (strongly) concave problems, wherein the objective function θ^{J} is replaced by an approximating function constructed at the current iteration $\left(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu \triangleq (\boldsymbol{\lambda}_i^\nu)_{i=1}^I\right) \in \Xi^J$, defined as

$$\widetilde{\theta}^{\text{J}}(\mathbf{x}, \boldsymbol{\lambda}\,;\, \mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \triangleq \sum_{i=1}^{I} \widetilde{\theta}_i^{\text{J}}(\mathbf{x}, \boldsymbol{\lambda}_i\,;\, \mathbf{x}^\nu, \boldsymbol{\lambda}_i^\nu) - \frac{\tau}{2}\left(\|\mathbf{x} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda} - \boldsymbol{\lambda}^\nu\|^2\right),$$

where $\tau > 0$ is a regularization constant, and for every $i = 1, \ldots, I$ each $\widetilde{\theta}_i^{\text{J}}$ is constructed by following the steps outlined next:

1. The concave term in (4.33) remains unchanged.

2. The convex term in (4.33) is replaced by a linear approximation. More precisely, for fixed i and for every $j \in \overline{\mathcal{J}}_i^{\text{cvx}} \cup \overline{\mathcal{J}}_i^{\text{cve}}$ we define the following linear approximation of $f_{i,j}$ at a given $\mathbf{x}^\nu \in \Xi$,

$$f_{i,j}(\mathbf{x}) \approx \widetilde{f}_{i,j}(\mathbf{x};\mathbf{x}^\nu) \triangleq f_{i,j}(\mathbf{x}^\nu) + \nabla f_{i,j}(\mathbf{x}^\nu)^T(\mathbf{x}-\mathbf{x}^\nu).$$

3. For the remaining term G_i notice that

$$G_i(\mathbf{x},\boldsymbol{\lambda}_i) = G_i(\mathbf{x}^\nu,\boldsymbol{\lambda}_i) + \left[G_i(\mathbf{x},\boldsymbol{\lambda}_i) - G_i(\mathbf{x}^\nu,\boldsymbol{\lambda}_i)\right],$$

   where the term in brackets can be approximated using a first order Taylor expansion at $(\mathbf{x}^\nu, \boldsymbol{\lambda}_i^\nu) \in \Xi^J$, that is

$$G_i(\mathbf{x},\boldsymbol{\lambda}_i) - G_i(\mathbf{x}^\nu,\boldsymbol{\lambda}_i) \approx \nabla_{\mathbf{x}} G_i(\mathbf{x}^\nu,\boldsymbol{\lambda}_i^\nu)^T(\mathbf{x}-\mathbf{x}^\nu).$$

   Hence, we approximate G_i as

$$G_i(\mathbf{x},\boldsymbol{\lambda}_i) \approx \widetilde{G}_i(\mathbf{x},\boldsymbol{\lambda}_i\,;\,\mathbf{x}^\nu,\boldsymbol{\lambda}^\nu) \triangleq G_i(\mathbf{x}^\nu,\boldsymbol{\lambda}_i) + \nabla_{\mathbf{x}} G_i(\mathbf{x}^\nu,\boldsymbol{\lambda}_i^\nu)^T(\mathbf{x}-\mathbf{x}^\nu).$$

From the three steps above, we define: for every $i = 1,\ldots,I$

$$\begin{aligned}\widetilde{\theta}_i^{\mathrm{J}}(\mathbf{x},\boldsymbol{\lambda}_i\,;\,\mathbf{x}^\nu,\boldsymbol{\lambda}_i^\nu) \triangleq & \left(\sum_{j\in\mathcal{J}_i^{\text{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}) + \sum_{j\in\mathcal{J}_i^{\text{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x})\right) \\ & + \left(\sum_{j\in\overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min}\,\widetilde{f}_{i,j}(\mathbf{x};\mathbf{x}^\nu) + \sum_{j\in\overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max}\,\widetilde{f}_{i,j}(\mathbf{x};\mathbf{x}^\nu)\right) \\ & + G_i(\mathbf{x}^\nu,\boldsymbol{\lambda}_i) + \nabla_{\mathbf{x}} G_i(\mathbf{x}^\nu,\boldsymbol{\lambda}_i^\nu)^T(\mathbf{x}-\mathbf{x}^\nu).\end{aligned} \tag{4.34}$$

An interesting feature of the approximation function above is that it exploits any "degree" of concavity present in every θ_i^J. For example, note that $\widetilde{G}_i$ preserves the concavity of $G_i(\mathbf{x},\bullet)$ on Λ_i for fixed $\mathbf{x} \in \Xi$, as opposed to classical approximation techniques that would simply linearize the function G_i in both variables, thus ignoring that level of concavity.

The description of the proposed iterative scheme is given in Algorithm 4.3.

We remark that a suitable practical termination criterion to be used in Step (`S.1`) of the mentioned algorithm is to stop iterating when $|\theta^{\mathrm{J}}(\mathbf{x}^{\nu}, \boldsymbol{\lambda}^{\nu}) - \theta^{\mathrm{J}}(\mathbf{x}^{\nu-1}, \boldsymbol{\lambda}^{\nu-1})| \leq \delta$ for some prescribed accuracy δ. The convergence properties of Algorithm 4.3 are stated formally in Proposition 4.6 below. It is worth mentioning that the proof of this proposition is an adaptation of the results in Chapter 3 to the $\mathrm{MSM}_{\mathrm{JO}}$ problem.

Algorithm 4.3: SCA-based Algorithm for the MSM Problem

Data: $\tau > 0$ and $(\mathbf{x}^0, \boldsymbol{\lambda}^0) \in \Xi^J$. Set $\nu = 0$.

(`S.0`): For every $i = 1, \ldots, I$ and $j = 1, \ldots, J$ compute $\rho_{i,j}^{\min}$ and $\rho_{i,j}^{\max}$.

(`S.1`): If $(\mathbf{x}^{\nu}, \boldsymbol{\lambda}^{\nu})$ satisfies a termination criterion, STOP.

(`S.2`): Compute

$$(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}) \triangleq \underset{\mathbf{x}\in\Xi,\, \boldsymbol{\lambda}\in\Lambda}{\operatorname{argmax}}\, \widetilde{\theta}^{\mathrm{J}}(\mathbf{x}, \boldsymbol{\lambda}\,;\, \mathbf{x}^{\nu}, \boldsymbol{\lambda}^{\nu}). \tag{4.35}$$

(`S.3`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

Proposition 4.6. Under assumptions A1-A5, for every initial point $(\mathbf{x}^0, \boldsymbol{\lambda}^0) \in \Xi^J$, the sequence $\{(\mathbf{x}^{\nu}, \boldsymbol{\lambda}^{\nu})\}$ produced by Algorithm 4.3 is well-defined. Moreover, if θ^J is bounded above on Ξ^J and the following two conditions hold:

(a) for every $i = 1, \ldots, I$ and $j = 1, \ldots, J$ there exists constants $L_{i,j} > 0$ such that $\|(h_{i,j}(\boldsymbol{\lambda}_i) - h_{i,j}(\widetilde{\boldsymbol{\lambda}}_i))\nabla f_{i,j}(\mathbf{x})\| \leq L_{i,j}\,\|\boldsymbol{\lambda}_i - \widetilde{\boldsymbol{\lambda}}_i\|$ for any $\mathbf{x} \in \Xi$ and $\boldsymbol{\lambda}_i, \widetilde{\boldsymbol{\lambda}}_i \in \Lambda_i$; and,

(b) the regularization constant is such that $\tau > 2\sum_{i=1}^{I}\sum_{j=1}^{J} L_{i,j}$;

then every accumulation point of $\{(\mathbf{x}^{\nu}, \boldsymbol{\lambda}^{\nu})\}$, if it exists, is a stationary solution of the joint optimization problem (4.29).

Proof. The first assertion of the proposition is clear.

For the second result, by induction, since $(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \in \Xi^J$, we have

$$\begin{aligned}
\theta^J(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) &= \widetilde{\theta}^J(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu\,;\, \mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \\
&\leq \widetilde{\theta}^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}\,;\, \mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \\
&= \sum_{i=1}^{I} \left(\sum_{j \in \mathcal{J}_i^{\text{cvx}}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}^{\nu+1}) + \sum_{j \in \mathcal{J}_i^{\text{cve}}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}^{\nu+1}) \right) \\
&\quad + \left(\sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min}\, \widetilde{f}_{i,j}(\mathbf{x}^{\nu+1}; \mathbf{x}^\nu) + \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max}\, \widetilde{f}_{i,j}(\mathbf{x}^{\nu+1}; \mathbf{x}^\nu) \right) \\
&\quad + \left(G_i(\mathbf{x}^\nu, \boldsymbol{\lambda}_i^{\nu+1}) + \nabla_{\mathbf{x}} G_i(\mathbf{x}^\nu, \boldsymbol{\lambda}_i^\nu)^T (\mathbf{x}^{\nu+1} - \mathbf{x}^\nu) \right) \\
&\quad - \frac{\tau}{2} \left(\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 \right).
\end{aligned}$$

It is not difficult to see that $\rho_{i,j}^{\min}\, \widetilde{f}_{i,j}(\mathbf{x}^{\nu+1}; \mathbf{x}^\nu) \leq \rho_{i,j}^{\min}\, f_{i,j}(\mathbf{x}^{\nu+1})$ and similarly $\rho_{i,j}^{\max}\, \widetilde{f}_{i,j}(\mathbf{x}^{\nu+1}; \mathbf{x}^\nu) \leq \rho_{i,j}^{\max}\, f_{i,j}(\mathbf{x}^{\nu+1})$ for every $j \in \overline{\mathcal{J}}_i^{\text{cvx}} \cup \overline{\mathcal{J}}_i^{\text{cve}}$. Applying this observation and the definition of G_i [cf. (4.7)] in the expression above, we obtain

$$\begin{aligned}
&\theta^J(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \\
&\leq \sum_{i=1}^{I} \left(\sum_{j: f_{i,j} \text{ cvx}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}^{\nu+1}) + \sum_{j: f_{i,j} \text{ cve}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}^{\nu+1}) \right) \\
&+ \left(\sum_{j: f_{i,j} \text{ cvx}} \left(h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) - \rho_{i,j}^{\min} \right) f_{i,j}(\mathbf{x}^\nu) + \sum_{j: f_{i,j} \text{ cve}} \left(\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) \right) (-f_{i,j}(\mathbf{x}^\nu)) \right) \\
&+ \left(\sum_{j: f_{i,j} \text{ cvx}} \left(h_{i,j}(\boldsymbol{\lambda}_i^\nu) - \rho_{i,j}^{\min} \right) \nabla f_{i,j}(\mathbf{x}^\nu) - \sum_{j: f_{i,j} \text{ cve}} \left(\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i^\nu) \right) \nabla f_{i,j}(\mathbf{x}^\nu) \right)^T (\mathbf{x}^{\nu+1} - \mathbf{x}^\nu) \\
&- \frac{\tau}{2} \left(\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 \right).
\end{aligned} \tag{4.36}$$

For every i and all j such that $f_{i,j}$ is convex, we add and subtract the terms

$$(h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) - \rho_{i,j}^{\min}) \nabla f_{i,j}(\mathbf{x}^\nu)^T (\mathbf{x}^{\nu+1} - \mathbf{x}^\nu)$$

in the right hand side of (4.36). Thus, it follows that

$$(h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) - \rho_{i,j}^{\min})f_{i,j}(\mathbf{x}^{\nu}) + (h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) - \rho_{i,j}^{\min})\nabla f_{i,j}(\mathbf{x}^{\nu})^T(\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu})$$
$$\leq (h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) - \rho_{i,j}^{\min})f_{i,j}(\mathbf{x}^{\nu+1}).$$

Similarly, for every i and all j such that $f_{i,j}$ is concave, we add and subtract the terms

$$(\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}))(-\nabla f_{i,j}(\mathbf{x}^{\nu}))^T(\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu})$$

obtaining

$$(\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}))(-f_{i,j}(\mathbf{x}^{\nu})) + (\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}))(-\nabla f_{i,j}(\mathbf{x}^{\nu}))^T(\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu})$$
$$\leq (\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}))(-f_{i,j}(\mathbf{x}^{\nu+1})).$$

Therefore, from (4.36), we obtain

$$\begin{aligned}
&\theta^J(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \\
&\leq \sum_{i=1}^{I} \left(\sum_{j:f_{i,j} \text{ cvx}} \rho_{i,j}^{\min} f_{i,j}(\mathbf{x}^{\nu+1}) + \sum_{j:f_{i,j} \text{ cve}} \rho_{i,j}^{\max} f_{i,j}(\mathbf{x}^{\nu+1}) \right) \\
&\quad + \sum_{j:f_{i,j} \text{ cvx}} [\, (h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) - \rho_{i,j}^{\min}) f_{i,j}(\mathbf{x}^{\nu+1}) \\
&\quad + (h_{i,j}(\boldsymbol{\lambda}_i^{\nu}) - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1})) \nabla f_{i,j}(\mathbf{x}^\nu)^T (\mathbf{x}^{\nu+1} - \mathbf{x}^\nu) \,] \\
&\quad \sum_{j:f_{i,j} \text{ cve}} [\, (\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}))(-f_{i,j}(\mathbf{x}^{\nu+1})) \\
&\quad + (h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) - h_{i,j}(\boldsymbol{\lambda}_i^{\nu}))(-\nabla f_{i,j}(\mathbf{x}^\nu))^T (\mathbf{x}^{\nu+1} - \mathbf{x}^\nu) \,] \\
&\quad - \frac{\tau}{2} \left(\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 \right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{I} \left(\sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1}) f_{i,j}(\mathbf{x}^{\nu+1}) \right) \\
&\quad + \left(\sum_{j=1}^{J} (h_{i,j}(\boldsymbol{\lambda}_i^{\nu}) - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1})) \nabla f_{i,j}(\mathbf{x}^\nu)^T (\mathbf{x}^{\nu+1} - \mathbf{x}^\nu) \right) \\
&\quad - \frac{\tau}{2} \left(\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 \right)
\end{aligned}$$

$$\begin{aligned}
&= \theta^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}) + \sum_{i=1}^{I} \sum_{j=1}^{J} (h_{i,j}(\boldsymbol{\lambda}_i^{\nu}) - h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1})) \nabla f_{i,j}(\mathbf{x}^\nu)^T (\mathbf{x}^{\nu+1} - \mathbf{x}^\nu) \\
&\quad - \frac{\tau}{2} \left(\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 \right).
\end{aligned}$$

Invoking Cauchy-Schwartz inequality along with assumption (a), from the expression above, it follows that

$$\begin{aligned}
\theta^J(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) &\leq \theta^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}) + \sum_{i=1}^{I} \sum_{j=1}^{J} L_{i,j} \|\boldsymbol{\lambda}_i^{\nu+1} - \boldsymbol{\lambda}_i^\nu\| \, \|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\| \\
&\quad - \frac{\tau}{2} \left(\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 \right) \\
&\leq \theta^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}) + \sum_{i=1}^{I} \sum_{j=1}^{J} L_{i,j} \left(\|\boldsymbol{\lambda}_i^{\nu+1} - \boldsymbol{\lambda}_i^\nu\|^2 + \|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 \right) \\
&\quad - \frac{\tau}{2} \left(\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 \right) \\
&\leq \theta^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}) + \sum_{i=1}^{I} \sum_{j=1}^{J} L_{i,j} \left(\|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 + \|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 \right) \\
&\quad - \frac{\tau}{2} \left(\|\mathbf{x}^{\nu+1} - \mathbf{x}^\nu\|^2 + \|\boldsymbol{\lambda}^{\nu+1} - \boldsymbol{\lambda}^\nu\|^2 \right) \\
&= \theta^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}) - \left(\frac{\tau}{2} - \sum_{i=1}^{I} \sum_{j=1}^{J} L_{i,j} \right) \|(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}) - (\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu)\|^2.
\end{aligned}$$

Thus, under the assumptions that θ^J is bounded above on Ξ^J and choosing $\tau > 2 \sum_{i=1}^{I} \sum_{j=1}^{J} L_{i,j}$, the sequence $\{(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu)\}$ converges and

$$\lim_{\nu \to \infty} \|(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}) - (\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu)\|^2 = 0.$$

By the variational principle, for every $\nu \geq 0$ and for all $(\mathbf{x}, \boldsymbol{\lambda}) \in \Xi^J$

$$\begin{aligned}
0 &\geq [(\mathbf{x}, \boldsymbol{\lambda}) - (\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1})]^T \, \nabla \widetilde{\theta}^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1} \, ; \, \mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \\
&= (\mathbf{x} - \mathbf{x}^{\nu+1})^T \, \nabla_{\mathbf{x}} \widetilde{\theta}^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1} \, ; \, \mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \\
&\quad + (\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\nu+1})^T \, \nabla_{\boldsymbol{\lambda}} \widetilde{\theta}^J(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1} \, ; \, \mathbf{x}^\nu, \boldsymbol{\lambda}^\nu)
\end{aligned} \tag{4.37}$$

where

$$
\begin{aligned}
&\nabla_{\mathbf{x}}\widetilde{\theta}^J(\mathbf{x}^{\nu+1},\boldsymbol{\lambda}^{\nu+1}\,;\,\mathbf{x}^\nu,\boldsymbol{\lambda}^\nu)\\
&=\sum_{i=1}^{I}\left(\sum_{j\in\mathcal{J}_i^{\text{cvx}}}\rho_{i,j}^{\min}\,\nabla f_{i,j}(\mathbf{x}^{\nu+1})+\sum_{j\in\mathcal{J}_i^{\text{cve}}}\rho_{i,j}^{\max}\,\nabla f_{i,j}(\mathbf{x}^{\nu+1})\right)\\
&\quad+\left(\sum_{j\in\overline{\mathcal{J}}_i^{\text{cvx}}}\rho_{i,j}^{\min}\,\nabla f_{i,j}(\mathbf{x}^{\nu})+\sum_{j\in\overline{\mathcal{J}}_i^{\text{cve}}}\rho_{i,j}^{\max}\,\nabla f_{i,j}(\mathbf{x}^{\nu})\right)\\
&\quad+\left(\sum_{j:f_{i,j}\ \text{cvx}}\big(h_{i,j}(\boldsymbol{\lambda}_i^\nu)-\rho_{i,j}^{\min}\big)\,\nabla f_{i,j}(\mathbf{x}^\nu)+\sum_{j:f_{i,j}\ \text{cve}}\big(\rho_{i,j}^{\max}-h_{i,j}(\boldsymbol{\lambda}_i^\nu)\big)\,(-\nabla f_{i,j}(\mathbf{x}^\nu))\right)\\
&\quad-\tau(\mathbf{x}^{\nu+1}-\mathbf{x}^\nu),
\end{aligned}
$$

and

$$
\begin{aligned}
&\nabla_{\boldsymbol{\lambda}}\widetilde{\theta}^J(\mathbf{x}^{\nu+1},\boldsymbol{\lambda}^{\nu+1}\,;\,\mathbf{x}^\nu,\boldsymbol{\lambda}^\nu)\\
&=\sum_{i=1}^{I}\left(\sum_{j:f_{i,j}\ \text{cvx}}\nabla h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1})f_{i,j}(\mathbf{x}^\nu)+\sum_{j:f_{i,j}\ \text{cve}}\nabla h_{i,j}(\boldsymbol{\lambda}_i^{\nu+1})f_{i,j}(\mathbf{x}^\nu)\right)-\tau(\boldsymbol{\lambda}^{\nu+1}-\boldsymbol{\lambda}^\nu).
\end{aligned}
$$

Finally, if $(\mathbf{x}^\infty,\boldsymbol{\lambda}^\infty)\in\Xi^J$ is an accumulation point of the sequence $\{(\mathbf{x}^\nu,\boldsymbol{\lambda}^\nu)\}$, then passing the limit along an appropriate subsequence in (4.37) we obtain: for all $(\mathbf{x},\boldsymbol{\lambda})\in\Xi^J$

$$
\begin{aligned}
0&\geq(\mathbf{x}-\mathbf{x}^\infty)^T\left[\sum_{i=1}^{I}\sum_{j=1}^{J}h_{i,j}(\boldsymbol{\lambda}_i^\infty)\,\nabla f_{i,j}(\mathbf{x}^\infty)\right]\\
&\quad+(\boldsymbol{\lambda}-\boldsymbol{\lambda}^\infty)^T\left[\sum_{i=1}^{I}\sum_{j=1}^{J}\nabla h_{i,j}(\boldsymbol{\lambda}_i^\infty)\,f_{i,j}(\mathbf{x}^\infty)\right]\\
&=(\mathbf{x}-\mathbf{x}^\infty)^T\nabla_{\mathbf{x}}\theta^J(\mathbf{x}^\infty,\boldsymbol{\lambda}^\infty)+(\boldsymbol{\lambda}-\boldsymbol{\lambda}^\infty)^T\nabla_{\boldsymbol{\lambda}}\theta^J(\mathbf{x}^\infty,\boldsymbol{\lambda}^\infty)\\
&=[(\mathbf{x},\boldsymbol{\lambda})-(\mathbf{x}^\infty,\boldsymbol{\lambda}^\infty)]^T\nabla\theta^J(\mathbf{x}^\infty,\boldsymbol{\lambda}^\infty)
\end{aligned}
$$

i.e. $(\mathbf{x}^\infty,\boldsymbol{\lambda}^\infty)$ is a stationary solution of (4.29). □

We remark that a sufficient condition guaranteeing assumption (a) of Proposition 4.6 is the Lipschitz continuity of the functions $h_{i,j}$ on Λ_i along with the uniform boundedness of the gradients of the functions $f_{i,j}$ over $\mathcal{X}$, for every $i = 1, \ldots, I$ and $j = 1, \ldots, J$. Of course, if each function $f_{i,j}$ is Lipschitz continuous on $\mathcal{X}$, the latter condition is readily satisfied. Note also that Algorithm 4.3 is shown to be convergent to stationary solutions of the MSM_{JO} problem, which under the conditions of Proposition 4.5(b) coincide with d-stationary solutions of the MSM problem.

4.4 Unwrapping the Algorithms

In the previous sections we have proposed two different approaches for attempting the computation of a solution for the MSM problem in (4.2). More precisely, the two approaches are: (i) a DC-based approach, discussed in Section 4.2, and (ii) a (smooth) Joint Optimization approach, developed in Section 4.3. These techniques gave rise to Algorithms 4.1 and 4.3, respectively. In this section, we explore in depth both algorithms and contrast them, highlighting their advantages and disadvantages if any.

An obvious difference between Algorithms 4.1 and 4.3 is that the former treats the variable $\boldsymbol{\lambda}$ implicitly, while the latter manages it explicitly. This fact translates into Algorithm 4.1 requiring the use of subgradients to treat the nondifferentiable term v [cf. (4.12)] of the DC-type function θ, while the smooth reformulation of the MSM problem translates into Algorithm 4.3 avoiding the use of subgradients. Hence, computationally speaking, Algorithm 4.1 requires the computation of I subgradients (see, Step 2 of the mentioned scheme), while Algorithm 4.3 increases the dimensions of the subproblems solved at each iteration by introducing I more variables and constraints (in comparison with the DC-based algorithm) into each of them. In order to contrast both schemes, let us unwrap them carefully by analyzing how the variables $\mathbf{x}^\nu$ and $\boldsymbol{\lambda}^\nu$ are updated at every iteration $\nu > 0$ for each approach.

(i) **DC-based Approach – Algorithm 4.1.** At any iteration $\nu > 0$ and given $\mathbf{x}^\nu \in \Xi$, Step 2 requires the computation of $\mu_{g_i}(\mathbf{x}^\nu) \in \partial g_i(\mathbf{x}^\nu)$ for every $i = 1, \ldots, I$. Hence, as previously discussed in Remark 4.3, this step needs the update of the variable $\boldsymbol{\lambda}^\nu$, which from (4.23), it is easily seen to be equivalent to: for every $i = 1, \ldots, I$

$$\boldsymbol{\lambda}_i^{\star,\nu} \in \underset{\boldsymbol{\lambda}_i \in \Lambda_i}{\operatorname{argmax}} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i) f_{i,j}(\mathbf{x}^\nu). \tag{4.38}$$

Then, after $\boldsymbol{\lambda}^\nu$ is updated accordingly to (4.38), Step 2 concludes by choosing any $\mu_{g_i}(\mathbf{x}^\nu) \in \partial g_i(\mathbf{x}^\nu)$ for every i where $\partial g_i(\mathbf{x}^\nu)$ is given in (4.19).

In Step 3, we set $\mathbf{x}^{\nu+1} \triangleq \widehat{\mathbf{x}}(\mathbf{x}^\nu)$. Thus, from (4.22), it follows easily that the variable $\mathbf{x}^\nu$ is updated by

$$\begin{aligned}\mathbf{x}^{\nu+1} \triangleq \underset{\mathbf{x} \in \Xi}{\operatorname{argmax}} \sum_{i=1}^{I} \Bigg(& u_i(\mathbf{x}) + \sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min} \nabla f_{i,j}(\mathbf{x}^\nu)^T \mathbf{x} \\ & + \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max} \nabla f_{i,j}(\mathbf{x}^\nu)^T \mathbf{x} + \mu_{g_i}(\mathbf{x}^\nu)^T \mathbf{x} \Bigg) - \frac{\tau}{2} \|\mathbf{x} - \mathbf{x}^\nu\|^2.\end{aligned} \tag{4.39}$$

(ii) **JO-based Approach – Algorithm 4.3.** In this case, the update of the variables $\mathbf{x}^\nu$ and $\boldsymbol{\lambda}^\nu$ is performed jointly in Step 2. However, notice that the optimization problem in (4.35) is separable in its respective variables. As a result, at any iteration $\nu > 0$ and given $(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu) \in \Xi^J$ the variables are updated accordingly to:

$$\boldsymbol{\lambda}_i^{\nu+1} \triangleq \underset{\boldsymbol{\lambda}_i \in \Lambda_i}{\operatorname{argmax}} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i) f_{i,j}(\mathbf{x}^\nu) - \frac{\tau}{2} \|\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_i^\nu\|^2, \tag{4.40}$$

and

$$\begin{aligned}\mathbf{x}^{\nu+1} \triangleq \underset{\mathbf{x} \in \Xi}{\operatorname{argmax}} \sum_{i=1}^{I} \Bigg(& u_i(\mathbf{x}) + \sum_{j \in \overline{\mathcal{J}}_i^{\text{cvx}}} \rho_{i,j}^{\min} \nabla f_{i,j}(\mathbf{x}^\nu)^T \mathbf{x} \\ & + \sum_{j \in \overline{\mathcal{J}}_i^{\text{cve}}} \rho_{i,j}^{\max} \nabla f_{i,j}(\mathbf{x}^\nu)^T \mathbf{x} + \nabla_{\mathbf{x}} G_i(\mathbf{x}^\nu, \boldsymbol{\lambda}_i^\nu)^T \mathbf{x} \Bigg) - \frac{\tau}{2} \|\mathbf{x} - \mathbf{x}^\nu\|^2.\end{aligned} \tag{4.41}$$

Contrasting the problems in (4.38) and (4.40), it is clear that the update of the variable $\boldsymbol{\lambda}^\nu$ differs only on the regularization term used in the JO-based approach. Thus, in the JO-based approach this update requires the solution of a strongly concave problem generating a unique choice for $\boldsymbol{\lambda}^{\nu+1}$. It is worth mentioning that, in the applications considered in Section 4.5 the problem in (4.38) simplifies to a scalar linear program, while (4.40) gives rise to a scalar quadratic problem, where both problems are solvable in closed form, and thus, these updates are computationally inexpensive.

With respect to the update of the variable $\mathbf{x}^\nu$, from (4.39) and (4.41) it follows that the only difference occurs in the last term in the summation of the objective functions of such problems. Such terms correspond to the approximations of the nondifferentiable (but convex) functions g_i in θ, and to the differentiable functions G_i in θ^J. Notice that, the DC-based scheme chooses any $\mu_{g_i}(\mathbf{x}^\nu) \in \partial g_i(\mathbf{x}^\nu)$ for every $i = 1, \ldots, I$ to linearize such terms, while the JO-based scheme uses the unique choice of the gradient (with respect to $\mathbf{x}$) of the function G_i to deal with them. Hence, roughly speaking, the update of $\mathbf{x}^\nu$ in the JO-based scheme can be viewed as a particular case of the one used in the DC-based algorithm. Finally, it is important to remark that in Algorithm 4.3 the pair $(\mathbf{x}^\nu, \boldsymbol{\lambda}^\nu)$ is updated simultaneously, while in Algorithm 4.1 those variables are update sequentially.

Let us turn our attention to the convergence properties of Algorithms 4.1 and 4.3. Notice that the conditions required for the convergence of Algorithm 4.3, stated in Proposition 4.6, are more stringent than those needed for the convergence of Algorithm 4.1 (refer to Proposition 4.3). In particular, the DC-based algorithm is shown to be convergent for any $\tau > 0$, while the JO-based scheme requires a particular choice of this regularization constant to ensure sufficient decrease (refer to Proposition 4.6, condition (b)). Thus, whenever this condition is difficult to evaluate or leads to "large" values of τ that may slow down the convergence of Algorithm 4.3, the DC scheme is always a choice. Refer to Chapter 3 for more details with regard to the choice of the parameters in the smooth joint optimization approach.

It is important to emphasize that Proposition 4.3 guarantees the convergence of Algorithm 4.1 to *critical points* of the $\mathrm{MSM}_{\mathrm{DC}}$ problem in (4.13), while Proposition 4.6 states that Algorithm 4.3 converges to *stationary solutions* of the $\mathrm{MSM}_{\mathrm{JO}}$ problem in (4.29). Note that, such first order points (stationary solutions or critical points) coincide with d-stationary solutions of the MSM problem in (4.2) if the conditions in the respective Propositions 4.2(d) or 4.5(b) are satisfied. Thus, it is clear that, given $\mathbf{x}^\star \in \Xi$ produced by either Algorithm 4.1 or 4.3, if the sets $\Lambda_i^\star(\mathbf{x}^\star)$ are a singleton for every $i = 1, \ldots, I$, then the concepts of stationary solutions, critical points and d-stationary solutions are all equivalent. Unfortunately, this is not always the case and the connections of the critical points and stationary solutions back to the d-stationary solutions of the original MSM problem are generally in jeopardy.

For the case where the functions $h_{i,j}$ are linear for every $i = 1, \ldots, I$ and all $j = 1, \ldots, J$ more can be said. In particular, the next proposition establishes that the concepts of critical points of the MSM_{DC} and stationary solutions of the MSM_{JO} are equivalent under this assumption. It is worth mentioning, that this case is of our particular interest, since in the physical layer based security applications discussed in Section 4.5 the functions $h_{i,j}$ are linear.

Proposition 4.7. Assume that the functions $h_{i,j}$ are linear for all $i = 1, \ldots, I$ and every $j = 1, \ldots, J$. The vector $\mathbf{x}^\star \in \Xi$ is a critical point of the MSM_{DC} problem in (4.13) if and only if there exists $\boldsymbol{\lambda}^\star \in \Lambda$ such that the pair $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ is a stationary solution of the MSM_{JO} problem in (4.29).

Proof. First, suppose that $\mathbf{x}^\star \in \Xi$ is a critical point of (4.13) i.e. $\mathbf{0} \in \nabla u(\mathbf{x}^\star) - \partial v(\mathbf{x}^\star) - \mathcal{N}_\Xi(\mathbf{x}^\star)$ if and only if, for every $i = 1, \ldots, I$, there exists $\mu_{g_i}(\mathbf{x}^\star) \in \partial g_i(\mathbf{x}^\star)$ such that: for all $\mathbf{x} \in \Xi$

$$\begin{aligned}\sum_{i=1}^{I} \Bigg(& \sum_{j: f_{i,j}\ \text{cvx}} \rho_{i,j}^{\min} \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) + \\ & \sum_{j: f_{i,j}\ \text{cve}} \rho_{i,j}^{\max} \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) + \mu_{g_i}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \Bigg) \leq 0.\end{aligned} \tag{4.42}$$

From (4.19), for every $i = 1, \ldots, I$, there must exist families $\{\boldsymbol{\lambda}_{i,\ell}\}_{\ell=1}^{L_i} \subset \Lambda_i^\star(\mathbf{x}^\star)$ and non-negative weights $\{w_{i,\ell}\}_{\ell=1}^{L_i}$ satisfying $\displaystyle\sum_{\ell=1}^{L_i} w_{i,\ell} = 1$, such that any element $\mu_{g_i}(\mathbf{x}^\star)$ of $\partial g_i(\mathbf{x}^\star)$ is of the form:

$$\mu_{g_i}(\mathbf{x}^\star) = \sum_{\ell=1}^{L_i} w_{i,\ell} \left(\sum_{j:f_{i,j}\ \text{cvx}} \left(h_{i,j}(\boldsymbol{\lambda}_{i,\ell}) - \rho_{i,j}^{\min}\right) \nabla f_{i,j}(\mathbf{x}^\star) + \sum_{j:f_{i,j}\ \text{cve}} \left(\rho_{i,j}^{\max} - h_{i,j}(\boldsymbol{\lambda}_{i,\ell})\right) \left(-\nabla f_{i,j}(\mathbf{x}^\star)\right) \right). \tag{4.43}$$

Replacing (4.43) for every $i = 1, \ldots, I$ in (4.42) and after some manipulations we obtain: for all $\mathbf{x} \in \Xi$

$$\begin{aligned} 0 &\geq \sum_{i=1}^{I} \left(\sum_{\ell=1}^{L_i} w_{i,\ell} \sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_{i,\ell}) \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \right) \\ &= \sum_{i=1}^{I} \left(\sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i^\star) \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) \right), \end{aligned} \tag{4.44}$$

where, in the last equality, we invoked the linearity of each function $h_{i,j}$ and set $\boldsymbol{\lambda}_i^\star \triangleq \sum_{\ell=1}^{L_i} w_{i,\ell}\, \boldsymbol{\lambda}_{i,\ell}$ for all $i = 1, \ldots, I$. It is clear that, for every i, $\boldsymbol{\lambda}_i^\star \in \Lambda_i^\star(\mathbf{x}^\star)$, hence:

$$\sum_{j=1}^{J} f_{i,j}(\mathbf{x}^\star)\, \nabla h_{i,j}(\boldsymbol{\lambda}_i^\star)^T (\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_i^\star) \leq 0 \qquad \forall\, \boldsymbol{\lambda}_i \in \Lambda_i. \tag{4.45}$$

Adding inequalities (4.44) and (4.45) over i, we obtain the desired result, that is, the pair $\left(\mathbf{x}^\star, \boldsymbol{\lambda}^\star \triangleq (\boldsymbol{\lambda}_i^\star)_{i=1}^I\right)$ is a stationary solution of the smooth maximization problem (4.29).

For the converse argument, suppose that the pair $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) \in \Xi^{\mathrm{J}}$ is a stationary solution (4.29). Thus, for all $\mathbf{x} \in \Xi$ and all $\boldsymbol{\lambda} \in \Lambda$ it holds that

$$\sum_{i=1}^{I} \left(\sum_{j=1}^{J} h_{i,j}(\boldsymbol{\lambda}_i^\star) \nabla f_{i,j}(\mathbf{x}^\star)^T (\mathbf{x} - \mathbf{x}^\star) + \sum_{j=1}^{J} f_{i,j}(\mathbf{x}^\star) \nabla h_{i,j}(\boldsymbol{\lambda}_i^\star)^T (\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_i^\star) \right) \leq 0.$$

Letting $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_i^\star$ for every $i = 1, \ldots, I$ in the inequality above the desired results follows readily. $\square$

Finally, from the overall analysis presented above, it follows that both

Algorithms 4.1 and 4.3 are expected to perform similarly in practice. In the next section, we apply such schemes to concrete resource allocation problems and test both of them numerically (refer to, Subsection 4.5.4).

4.5 Applications in Physical Layer Based Security

In this section, we apply the methods developed in Sections 4.2 and 4.3 to resource allocation problems in the context of physical layer based security. In the applications below, we consider a SISO system, i.e. a communication network with transmitter-receiver pairs' equipped with a single antenna, where the spectrum is managed dynamically. Based on the secrecy rate concept (see, e.g., [47]), the main problem is to allocate the power budget of the network entities so that their transmissions are kept secure. We study this problem under two different assumptions: first, we deal with the case where the users communicate over multiple non-orthogonal subchannels (see, Subsection 4.5.1); and second, we consider OFDMA transmissions (see, Subsection 4.5.2). For the first setting, we model the system design as a single optimization problem, leading to centralized algorithms; while for the second case, we formulate the system design as a noncooperative game [76] leading to distributed schemes. In Subsections 4.5.3, we present some possible extensions of the aforementioned problems. Subsection 4.5.4 presents some numerical experiments.

4.5.1 SISO Secrecy Sum-Rate Maximization Problem

4.5.1.1 System Model and Notation

We consider a wireless communication system composed of Q transmitter-receiver pairs (denoted by $q = 1, \ldots, Q$) which correspond to the legitimate users of the system, J friendly jammers (denoted by $j = 1, \ldots, J$) and a single eavesdropper (denoted by e); see Figure 4.3. In this setting, each legitimate user's transmitter wants to communicate (in a secure way) with its corresponding legitimate receiver over a set of N parallel subchannels (denoted by $k = 1, \ldots, N$). The friendly jammers are entities willing to cooperate with the legitimate parties by introducing judicious interference such as to impair the eavesdroppers ability to decode the messages between intended nodes.

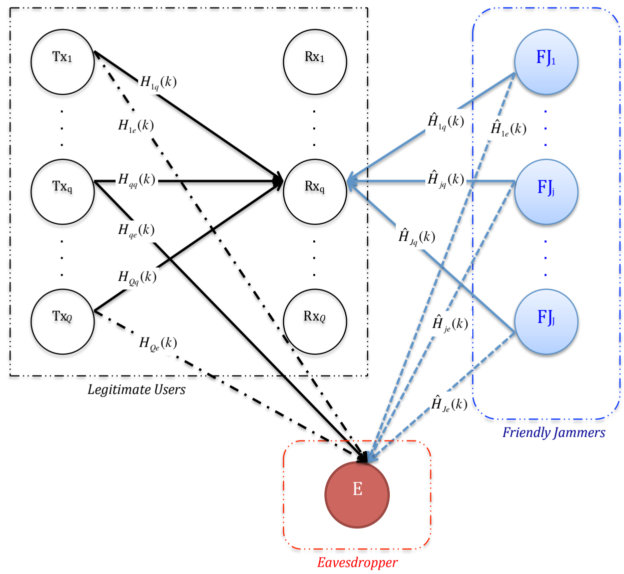

Figure 4.3: System model composed of the legitimate users, the friendly jammers, and one eavesdropper. The arrows illustrate the interference perceived by the q-th user and the eavesdropper over the k-th subchannel.

In the following discussion, let:

- $p_q(k)$ denotes the power allocation of transmitter q over subchannel k.
- $\widehat{p}_j(k)$ denotes the power allocation of the friendly jammer j over subchannel k.
- $P_q^{\max}$ denotes the power budget of the q-th transmitter.
- $\widehat{P}_j^{\max}$ denotes the power budget of the j-th friendly jammer.
- $H_{qq}(k)$ represents the channel gain between the transmitter of user q and its corresponding receiver over subchannel k. Similarly, $H_{qe}(k)$ represents the channel gain coefficient for the k-th subchannel between the transmitter of user q and the eavesdropper.
- $H_{rq}(k)$ represents the channel gain between the transmitter of user r and the receiver of user q over subchannel k. Similarly, $H_{re}(k)$ represents the channel gain coefficient for the k-th subchannel between the transmitter of user r and the eavesdropper.

- $\widehat{H}_{jq}(k)$ represents the channel gain between the transmitter of the friendly jammer j and the receiver of user q along subchannel k. Likewise, $\widehat{H}_{je}(k)$ represents the channel gain coefficient for the k-th subchannel between the transmitter of the friendly jammer j and the eavesdropper.

- $\sigma_q^2(k)$ denotes the variance of the noise at the receiver of user q along subchannel k.

We assume that Channel State Information (CSI) of the eavesdropper's (cross-)channels is available. We remark that this is a common assumption in the physical layer based security literature; see, e.g., [47, 24, 39, 59, 115, 104, 119, 36, 60]. There are practical situations in which the CSI can be obtained, see Chapter 3 - Section 3.5 and the references therein for some examples.

Under basic information theoretical assumptions, the maximum achievable rate on link q over subchannel k for a given power profile, defined by $\left(\mathbf{p} \triangleq (\mathbf{p}_q \triangleq (p_q(k))_{k=1}^N)_{q=1}^Q,\ \widehat{\mathbf{p}} \triangleq (\widehat{\mathbf{p}}_j \triangleq (\widehat{p}_j(k))_{k=1}^N)_{j=1}^J\right)$, is given by [22]

$$r_{qqk}(\mathbf{p}, \widehat{\mathbf{p}}) \triangleq \log\left(1 + \frac{H_{qq}(k)\, p_q(k)}{\sigma_q^2(k) + \sum_{r \neq q} H_{rq}(k)\, p_r(k) + \sum_{j=1}^J \widehat{H}_{jq}(k)\, \widehat{p}_j(k)}\right).$$

Similarly, the rate on the link between the transmitter of user q and the eavesdropper's receiver along subchannel k is

$$r_{qek}(\mathbf{p}, \widehat{\mathbf{p}}) \triangleq \log\left(1 + \frac{H_{qe}(k)\, p_q(k)}{\sigma_q^2(k) + \sum_{r \neq q} H_{re}(k)\, p_r(k) + \sum_{j=1}^J \widehat{H}_{je}(k)\, \widehat{p}_j(k)}\right).$$

Then, the *secrecy rate* on link q along subchannel k is (see, e.g., [47])

$$r_{qk}^s(\mathbf{p}, \widehat{\mathbf{p}}) \triangleq \left[r_{qqk}(\mathbf{p}, \widehat{\mathbf{p}}) - r_{qek}(\mathbf{p}, \widehat{\mathbf{p}})\right]^+, \tag{4.46}$$

where $[\bullet]^+$ denotes the Euclidean projection onto $\mathbb{R}_+$ i.e. $[x]^+ \triangleq \max(0, x)$.

4.5.1.2 Problem Formulation

We formulate the system design as a single optimization problem that seeks to maximize the system's utility, which we take as the *secrecy sum-rate* (denoted by r^s) i.e. the sum of the secrecy rate of all users along all subchannel. Assuming that each transmitter q has a limited power budget, that is $\sum_{k=1}^{N} p_q(k) \leq P_q^{\max}$, and likewise each jammer j i.e. $\sum_{k'=1}^{N} \widehat{p}_j(k') \leq \widehat{P}_j^{\max}$, then the system design corresponds to the following maximization problem:

$$\begin{aligned}
&\underset{(\mathbf{p},\widehat{\mathbf{p}})\geq \mathbf{0}}{\text{maximize}}\ \ r^s(\mathbf{p},\widehat{\mathbf{p}}) \triangleq \sum_{q=1}^{Q}\sum_{k=1}^{N}\ r^s_{qk}\left(\mathbf{p},\widehat{\mathbf{p}}\right) \\
&\ \text{subject to:} \\
&\qquad\qquad \sum_{k=1}^{N} p_q(k) \leq P_q^{\max} \quad\ \ \forall\, q=1,\ldots,Q \quad \textit{(private constraints)} \\
&\qquad\qquad \sum_{k'=1}^{N} \widehat{p}_j(k') \leq \widehat{P}_j^{\max} \quad\ \ \forall\, j=1,\ldots,J \quad \textit{(coupling constraints).}
\end{aligned} \tag{4.47}$$

For the sake of simplicity, let the polyhedral feasible set of (4.47) be denoted by

$$\mathcal{P} \triangleq \left\{ (\mathbf{p},\widehat{\mathbf{p}}) \geq \mathbf{0}\ :\ \sum_{k=1}^{N} p_q(k) \leq P_q^{\max} \quad \forall\, q \quad \text{and} \quad \sum_{k'=1}^{N} \widehat{p}_j(k') \leq \widehat{P}_j^{\max} \quad \forall\, j \right\}.$$

Notice that, for every $q = 1,\ldots,Q$ and every $k = 1,\ldots,N$, r^s_{qk} [cf. (4.46)] can be rewritten as

$$r^s_{qk}(\mathbf{p},\widehat{\mathbf{p}}) = \left[f_{qk,1}(\mathbf{p},\widehat{\mathbf{p}}) + f_{qk,2}(\mathbf{p},\widehat{\mathbf{p}})\right]^+, \tag{4.48}$$

where

$$\begin{aligned}
f_{qk,1}(\mathbf{p},\widehat{\mathbf{p}}) \triangleq{}& \log\left(\sigma_q^2(k) + \sum_{r=1}^{Q} H_{rq}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j(k)\right) \\
&+ \log\left(\sigma_q^2(k) + \sum_{r\neq q} H_{re}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j(k)\right)
\end{aligned} \tag{4.49}$$

$$\begin{aligned}
f_{qk,2}(\mathbf{p},\widehat{\mathbf{p}}) \triangleq{}& -\log\left(\sigma_q^2(k) + \sum_{r=1}^{Q} H_{re}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j(k)\right) \\
&- \log\left(\sigma_q^2(k) + \sum_{r\neq q} H_{rq}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j(k)\right).
\end{aligned} \tag{4.50}$$

Clearly, each $f_{qk,1}$ is a concave function and each $f_{qk,2}$ is a convex function. Let us rewrite the secrecy sum-rate maximization problem in (4.47) as follows:

$$\underset{(\mathbf{p},\widehat{\mathbf{p}})\in\mathcal{P}}{\text{maximize}} \sum_{q=1}^{Q}\sum_{k=1}^{N} \underset{\lambda_{qk}\in[0,1]}{\text{maximum}} \left(\lambda_{qk}\, f_{qk,1}(\mathbf{p},\widehat{\mathbf{p}}) + \lambda_{qk}\, f_{qk,2}(\mathbf{p},\widehat{\mathbf{p}})\right), \tag{4.51}$$

where each *discrete* max-function $[\bullet]^+$ in (4.48) is replaced by a *continuous* equivalent form. It is important to highlight the following three key observations about the SISO secrecy sum-rate maximization reformulation in (4.51):

i) It is not difficult to show that the optimizations problems in (4.47) and (4.51) are equivalent in terms of globally optimal solutions and d-stationary points;

ii) Furthermore, the maximization problem (4.51) is an instance of the MSM problem introduced in (4.2); and,

iii) It is easy to verify that assumptions A1-A6 are readily satisfied by the application in consideration.

Thus, capitalizing on these observations, we can apply either Algorithm 4.1 to find critical points of the DC-reformulation of (4.51) or Algorithm 4.3 to find stationary solutions of the joint optimization reformulation of such a problem. In what follows, we customize both schemes to the aforementioned problem.

4.5.1.3 DC-Programming Approach

Following the ideas of Section 4.2, lets start by finding an equivalent DC-decomposition of the secrecy sum-rate r^s. For that, it suffices to focus on each of its summands

$$r^s_{qk}(\mathbf{p},\widehat{\mathbf{p}}) = \underset{\lambda_{qk}\in[0,1]}{\text{maximum}} \left(\lambda_{qk}\, f_{qk,1}(\mathbf{p},\widehat{\mathbf{p}}) + \lambda_{qk}\, f_{qk,2}(\mathbf{p},\widehat{\mathbf{p}})\right).$$

Notice that, for this application, the constants defined in equation (4.4) become $\rho^{\max}_{qk} = 1$ and $\rho^{\min}_{qk} = 0$ for every $q = 1,\ldots,Q$ and every $k = 1,\ldots,N$.

Therefore, the expression in (4.5) simplifies to: for every $q = 1, \ldots, Q$ and $k = 1, \ldots, N$

$$r^s_{qk}(\mathbf{p}, \widehat{\mathbf{p}}) = f_{qk,1}(\mathbf{p}, \widehat{\mathbf{p}}) + \underset{\lambda_{qk} \in [0,1]}{\text{maximum}} \, [\lambda_{qk} f_{qk,2}(\mathbf{p}, \widehat{\mathbf{p}}) + (\lambda_{qk} - 1) f_{qk,1}(\mathbf{p}, \widehat{\mathbf{p}})]. \quad (4.52)$$

From the equation above, it is clear that the expression in (4.6) simplifies to the particular form

$$g_{qk}(\mathbf{p}, \widehat{\mathbf{p}}) \triangleq \underset{\lambda_{qk} \in [0,1]}{\text{maximum}} \; G_{qk}(\mathbf{p}, \widehat{\mathbf{p}}, \lambda_{qk}), \quad (4.53)$$

where

$$G_{qk}(\mathbf{p}, \widehat{\mathbf{p}}, \lambda_{qk}) \triangleq \lambda_{qk} \, f_{qk,2}(\mathbf{p}, \widehat{\mathbf{p}}) + (\lambda_{qk} - 1) \, f_{qk,1}(\mathbf{p}, \widehat{\mathbf{p}}).$$

From this discussion, the desired DC-decomposition of r^s follows readily. We highlight that this nontrivial decomposition will also lead us to derive iterative algorithms for more complex system designs as those studied in Subsection 4.5.3.

Under the observations above, Algorithm 4.4 specializes Algorithm 4.1 to the SISO secrecy sum-rate maximization problem. For notational convenience, when needed, we will denote each tuple $(\mathbf{p}, \widehat{\mathbf{p}})$ by $\mathbf{x} \triangleq (\mathbf{p}, \widehat{\mathbf{p}})$. For any $\tau > 0$, Algorithm 4.4 converges (in the sense of Proposition 4.3) to a critical point of the DC-reformulation of (4.51), which is simply:

$$\underset{(\mathbf{p}, \widehat{\mathbf{p}}) \in \mathcal{P}}{\text{maximize}} \;\; \sum_{q=1}^{Q} \sum_{k=1}^{N} f_{qk,1}(\mathbf{p}, \widehat{\mathbf{p}}) - (-g_{qk}(\mathbf{p}, \widehat{\mathbf{p}})) \,. \quad (4.54)$$

We remark that the feasible set $\mathcal{P}$ of the problem above is bounded, consequently the existence of an accumulation point of the sequence $\{\mathbf{x}^\nu\}$ produced by Algorithm 4.4 is guaranteed as required by the general convergence result in Proposition 4.3.

Let us highlight two aspects regarding the computation of a subgradient of g_{qk} [cf. (4.53)]. First, notice that the computation of (4.23), which corresponds to step (`S.2`) of Algorithm 4.4, is computationally inexpensive for this particular case. Namely, it only requires the solution of a scalar linear problem which reduces to the closed form expression in (4.55). Second, step (`S.3`) requires the computation of $\nabla_{\mathbf{x}} G_{qk}(\mathbf{x}, \lambda_{qk})$, which by the differentia-

Algorithm 4.4: DC-based Algorithm for the SISO Secrecy Sum-Rate Maximization Problem

Data: $\tau > 0$ and $\mathbf{x}^0 \triangleq (\mathbf{p}^0, \widehat{\mathbf{p}}^0) \in \mathcal{P}$. Set $\nu = 0$.

(`S.1`): If $\mathbf{x}^\nu \triangleq (\mathbf{p}^\nu, \widehat{\mathbf{p}}^\nu)$ satisfies a termination criterion, STOP.

(`S.2`): For $q = 1, \ldots, Q$ and $k = 1, \ldots, N$ compute

$$\lambda_{qk}^{\star,\nu} \in \begin{cases} \{0\} & \text{if} \quad f_{qk,1}(\mathbf{x}^\nu) + f_{qk,2}(\mathbf{x}^\nu) < 0 \\ \{1\} & \text{if} \quad f_{qk,1}(\mathbf{x}^\nu) + f_{qk,2}(\mathbf{x}^\nu) > 0 \\ [0,1] & \text{if} \quad f_{qk,1}(\mathbf{x}^\nu) + f_{qk,2}(\mathbf{x}^\nu) = 0. \end{cases} \tag{4.55}$$

(`S.3`): For $q = 1, \ldots, Q$ and $k = 1, \ldots, N$ compute

$$\mu_{g_{qk}}(\mathbf{x}^\nu) = \begin{cases} -\nabla f_{qk,1}(\mathbf{x}^\nu) & \text{if} \quad \lambda_{qk}^{\star,\nu} = 0 \\ \nabla f_{qk,2}(\mathbf{x}^\nu) & \text{if} \quad \lambda_{qk}^{\star,\nu} = 1 \\ \lambda_{qk}^{\star,\nu}\, \nabla f_{qk,2}(\mathbf{x}^\nu) + (\lambda_{qk}^{\star,\nu} - 1)\, \nabla f_{qk,1}(\mathbf{x}^\nu) & \text{if} \quad \lambda_{qk}^{\star,\nu} \in (0,1). \end{cases} \tag{4.56}$$

(`S.4`): Compute

$$\mathbf{x}^{\nu+1} \triangleq \underset{\mathbf{x}\in\mathcal{P}}{\operatorname{argmax}} \sum_{q=1}^{Q} \sum_{k=1}^{N} \left(f_{qk,1}(\mathbf{x}) + \mu_{g_{qk}}(\mathbf{x}^\nu)^T \mathbf{x} \right) - \frac{\tau}{2} \|\mathbf{x} - \mathbf{x}^\nu\|^2. \tag{4.57}$$

(`S.5`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

bility of each function $f_{qk,1}(\mathbf{x})$ and $f_{qk,2}(\mathbf{x})$, is given by: for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$

$$\nabla_{\mathbf{x}} G_{qk}(\mathbf{x}, \lambda_{qk}) = \lambda_{qk}\, \nabla f_{qk,2}(\mathbf{x}) + (\lambda_{qk} - 1)\, \nabla f_{qk,1}(\mathbf{x}), \tag{4.58}$$

where

$$\nabla f_{qk,1}(\mathbf{x}) \triangleq \begin{pmatrix} \left(\nabla_{\mathbf{p}_q} f_{qk,1}(\mathbf{x}) \triangleq \left(\frac{\partial f_{qk,1}(\mathbf{x})}{\partial p_q(k)} \right)_{k=1}^{N} \right)_{q=1}^{Q} \\ \left(\nabla_{\widehat{\mathbf{p}}_j} f_{qk,1}(\mathbf{x}) \triangleq \left(\frac{\partial f_{qk,1}(\mathbf{x})}{\partial \widehat{p}_j(k)} \right)_{k=1}^{N} \right)_{j=1}^{J} \end{pmatrix} \in \mathbb{R}^{N(Q+J)} \tag{4.59}$$

$$\nabla f_{qk,2}(\mathbf{x}) \triangleq \begin{pmatrix} \left(\nabla_{\mathbf{p}_q} f_{qk,2}(\mathbf{x}) \triangleq \left(\frac{\partial f_{qk,2}(\mathbf{x})}{\partial p_q(k)} \right)_{k=1}^{N} \right)_{q=1}^{Q} \\ \left(\nabla_{\widehat{\mathbf{p}}_j} f_{qk,2}(\mathbf{x}) \triangleq \left(\frac{\partial f_{qk,2}(\mathbf{x})}{\partial \widehat{p}_j(k)} \right)_{k=1}^{N} \right)_{j=1}^{J} \end{pmatrix} \in \mathbb{R}^{N(Q+J)}, \tag{4.60}$$

and the corresponding partial derivatives are given by:

$$\frac{\partial f_{qk,1}(\mathbf{x})}{\partial p_q(k)} = \frac{H_{qq}(k)}{\sigma_q^2(k) + \sum_{r=1}^{Q} H_{rq}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j(k)}$$

$$\begin{aligned}\frac{\partial f_{qk,1}(\mathbf{x})}{\partial p_r(k)} &= \frac{H_{rq}(k)}{\sigma_q^2(k) + \sum_{r=1}^{Q} H_{rq}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j(k)} \\ &\quad + \frac{H_{re}(k)}{\sigma_q^2(k) + \sum_{r\neq q} H_{re}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j(k)} \quad (\forall\, r \neq q)\end{aligned}$$

$$\begin{aligned}\frac{\partial f_{qk,1}(\mathbf{x})}{\partial \widehat{p}_j(k)} &= \frac{\widehat{H}_{jq}(k)}{\sigma_q^2(k) + \sum_{r=1}^{Q} H_{rq}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j(k)} \\ &\quad + \frac{\widehat{H}_{je}(k)}{\sigma_q^2(k) + \sum_{r\neq q} H_{re}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j(k)} \quad (\forall\, j = 1, \ldots, J).\end{aligned}$$

Similarly,

$$\frac{\partial f_{qk,2}(\mathbf{x})}{\partial p_q(k)} = \frac{-H_{qe}(k)}{\sigma_q^2(k) + \sum_{r=1}^{Q} H_{re}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j(k)}$$

$$\begin{aligned}\frac{\partial f_{qk,2}(\mathbf{x})}{\partial p_r(k)} &= \frac{-H_{re}(k)}{\sigma_q^2(k) + \sum_{r=1}^{Q} H_{re}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j(k)} \\ &- \frac{H_{rq}(k)}{\sigma_q^2(k) + \sum_{r\neq q} H_{rq}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j(k)} \quad (\forall\, r \neq q)\end{aligned}$$

$$\begin{aligned}\frac{\partial f_{qk,2}(\mathbf{x})}{\partial \widehat{p}_j(k)} &= \frac{-\widehat{H}_{je}(k)}{\sigma_q^2(k) + \sum_{r=1}^{Q} H_{re}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j(k)} \\ &- \frac{\widehat{H}_{jq}(k)}{\sigma_q^2(k) + \sum_{r\neq q} H_{rq}(k)\, p_r(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j(k)} \quad (\forall\, j = 1, \ldots, J).\end{aligned}$$

Notice that Algorithm 4.4 is centralized i.e. it requires a level of coordination among the entities in the network. An example of a natural application of this scheme occurs in broadband wireless networks, where the base station can act as the central authority. In this case, the base station is capable of estimating the users' channels. After the CSI is available, the base station executes Algorithm 4.4 and assigns the resources, both the power and subchannels, in accordance to the output obtained form the scheme.

To conclude this section, we examine the relation between the critical points of (4.54) (produced by Algorithm 4.4) and the d-stationary solutions of the maximization problem (4.47). Recall that, from Proposition 4.2(c) it follows that every d-stationary solution of the secrecy sum-rate maximization problem is a critical point of its DC-reformulation. However, the converse is not always true since the conditions of Proposition 4.2(d) are not satisfied by our application. Therefore, in the following discussion, we exploit the particular structure of our problem in order to derive some results connecting those concepts. For that, let us start by calculating the directional derivative of $r^s(\mathbf{x})$ along any direction $\mathbf{d} \in \mathbb{R}^n$, i.e.

$$r^{s\prime}(\mathbf{x};\mathbf{d}) = \sum_{q=1}^{Q}\sum_{k=1}^{N} r_{qk}^{s\prime}(\mathbf{x};\mathbf{d}) \tag{4.61}$$

where from [13, Eq. 2.124] we have that: for all $q = 1,\ldots,Q$ and $k = 1,\ldots,N$

$$r_{qk}^{s\prime}(\mathbf{x};\mathbf{d}) \triangleq \begin{cases} 0 & \text{if } f_{qk,1}(\mathbf{x}) + f_{qk,2}(\mathbf{x}) < 0 \\ \left(\nabla f_{qk,1}(\mathbf{x}) + \nabla f_{qk,2}(\mathbf{x})\right)^T \mathbf{d} & \text{if } f_{qk,1}(\mathbf{x}) + f_{qk,2}(\mathbf{x}) > 0 \\ \left(\left(\nabla f_{qk,1}(\mathbf{x}) + \nabla f_{qk,2}(\mathbf{x})\right)^T \mathbf{d}\right)^+ & \text{if } f_{qk,1}(\mathbf{x}) + f_{qk,2}(\mathbf{x}) = 0. \end{cases} \tag{4.62}$$

Suppose that $\mathbf{x}^\star$ is a critical point of the DC program (4.54) produced by Algorithm 4.4, i.e.

$$\mathbf{0} \in \sum_{q=1}^{Q}\sum_{k=1}^{N} \nabla f_{qk,1}(\mathbf{x}^\star) + \partial g_{qk}(\mathbf{x}^\star) - \mathcal{N}_{\mathcal{P}}(\mathbf{x}^\star). \tag{4.63}$$

Lets evaluate (4.63) element-wise, that is, for any pair (q,k) the following three cases can happen at a critical point $\mathbf{x}^\star$.

- **Case I** $-$ $f_{qk,1}(\mathbf{x}^\star) + f_{qk,2}(\mathbf{x}^\star) < 0$.

  In this case $\mu_{g_{qk}}(\mathbf{x}^\star) = -\nabla f_{qk,1}(\mathbf{x}^\star)$, thus for all $\mathbf{x} \in \mathcal{P}$

  $$\left(\nabla f_{qk,1}(\mathbf{x}^\star) + \mu_{g_{qk}}(\mathbf{x}^\star)\right)^T (\mathbf{x} - \mathbf{x}^\star) = 0 = r_{qk}^{s\prime}(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star),$$

  where in the last equality we invoked (4.62).

- **Case II** $-$ $f_{qk,1}(\mathbf{x}^\star) + f_{qk,2}(\mathbf{x}^\star) > 0$.

  In this case $\mu_{g_{qk}}(\mathbf{x}^\star) = \nabla f_{qk,2}(\mathbf{x}^\star)$, thus for all $\mathbf{x} \in \mathcal{P}$

  $$\left(\nabla f_{qk,1}(\mathbf{x}^\star) + \mu_{g_{qk}}(\mathbf{x}^\star)\right)^T (\mathbf{x} - \mathbf{x}^\star) = r_{qk}^{s\prime}(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star),$$

  where, again, the equality above follows from (4.62).

- **Case III** $-$ $f_{qk,1}(\mathbf{x}^\star) + f_{qk,2}(\mathbf{x}^\star) = 0$.

As expected, this is the case that puts in jeopardy the equivalence between a critical point and a d-stationary solution of the respective problems. Note that, under this setting, in Step `(S.3)` of Algorithm 4.4 the subgradient of g_{qk} at $\mathbf{x}^\star$ can be any

$$\mu_{g_{qk}}(\mathbf{x}^\star) \in \left\{ \lambda_{qk}^\star \, \nabla f_{qk,2}(\mathbf{x}^\star) + (1-\lambda_{qk}^\star)\left(-\nabla f_{qk,1}(\mathbf{x}^\star)\right) \,:\, \lambda_{qk}^\star \in [0,1] \right\}.$$

Besides, it is not difficult to see that $f_{qk,1}(\mathbf{x}^\star) + f_{qk,2}(\mathbf{x}^\star) = 0$ if and only if either $p_q^\star(k) = 0$ or

$$\frac{H_{qq}(k)}{\sigma_q^2(k) + \sum_{r\neq q} H_{rq}(k)\, p_r^\star(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j^\star(k)} = \frac{H_{qe}(k)}{\sigma_q^2(k) + \sum_{r\neq q} H_{re}(k)\, p_r^\star(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j^\star(k)},$$

which is equivalent to

$$M_{qk}(\mathbf{x}^\star) = 0$$

where, for notational simplicity, we define

$$\begin{aligned} M_{qk}(\mathbf{x}^\star) \triangleq & \sum_{r\neq q} \left(H_{qq}(k)H_{re}(k) - H_{qe}(k)H_{rq}(k)\right) p_r^\star(k) \\ & + \sum_{j=1}^{J} \left(H_{qq}(k)\widehat{H}_{je}(k) - H_{qe}(k)\widehat{H}_{jq}(k)\right) \widehat{p}_j^\star(k) \\ & + \left(H_{qq}(k) - H_{qe}(k)\right) \sigma_q^2(k). \end{aligned} \tag{4.64}$$

This observation lead us to consider the following four possible situations.

(a) If $p_q^\star(k) = 0$ and $M_{qk}(\mathbf{x}^\star) = 0$, using the expressions in (4.59) and (4.60), it is not difficult to show that: for all $\mathbf{x} \in \mathcal{P}$

$$\left(\nabla f_{qk,1}(\mathbf{x}^\star) + \nabla f_{qk,2}(\mathbf{x}^\star)\right)^T (\mathbf{x} - \mathbf{x}^\star) = 0.$$

Therefore, from (4.62) it follows that: for all $\mathbf{x} \in \mathcal{P}$

$$r_{qk}^{s\,\prime}(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star) = \left(\nabla f_{qk,1}(\mathbf{x}^\star) + \nabla f_{qk,2}(\mathbf{x}^\star)\right)^T (\mathbf{x} - \mathbf{x}^\star) = 0.$$

Note that, if $\lambda_{qk}^{\star} = 1$ (or $\lambda_{qk}^{\star} = 0$) then the subgradient is $\mu_{g_{qk}}(\mathbf{x}^{\star}) = \nabla f_{qk,2}(\mathbf{x}^{\star})$ (or $\mu_{g_{qk}}(\mathbf{x}^{\star}) = -\nabla f_{qk,1}(\mathbf{x}^{\star})$); and, from the equation above, we obtain the desired correspondence between the value of the (q,k)-th directional derivative and that of the respective term $\left(\nabla f_{qk,1}(\mathbf{x}^{\star}) + \mu_{g_{qk}}(\mathbf{x}^{\star})\right)^T (\mathbf{x} - \mathbf{x}^{\star})$ of (4.63).

(b) If $p_q^{\star}(k) = 0$ and $M_{qk}(\mathbf{x}^{\star}) > 0$, from the expressions in (4.59) and (4.60), it is easy to show that

$$\left(\nabla f_{qk,1}(\mathbf{x}^{\star}) + \nabla f_{qk,2}(\mathbf{x}^{\star})\right)^T (\mathbf{x} - \mathbf{x}^{\star}) = \left(\frac{\partial f_{qk,1}(\mathbf{x}^{\star})}{\partial p_q(k)} - \frac{\partial f_{qk,2}(\mathbf{x}^{\star})}{\partial p_q(k)}\right) p_q(k) \geq 0$$

for all $p_q(k) \geq 0$. Hence $\left(\nabla f_{qk,1}(\mathbf{x}^{\star}) + \nabla f_{qk,2}(\mathbf{x}^{\star})\right)^T (\mathbf{x} - \mathbf{x}^{\star}) \geq 0$ for all $\mathbf{x} \in \mathcal{P}$. As a result, from (4.62) it is clear that: for all $\mathbf{x} \in \mathcal{P}$

$$r_{qk}^{s\prime}(\mathbf{x}^{\star}; \mathbf{x} - \mathbf{x}^{\star}) = \left(\nabla f_{qk,1}(\mathbf{x}^{\star}) + \nabla f_{qk,2}(\mathbf{x}^{\star})\right)^T (\mathbf{x} - \mathbf{x}^{\star}) \geq 0.$$

Note that, if $\lambda_{qk}^{\star} = 1$ then $\mu_{g_{qk}}(\mathbf{x}^{\star}) = \nabla f_{qk,2}(\mathbf{x}^{\star})$, and from the equation above we obtain the desired correspondence between the value of the (q,k)-th directional derivative and that of the term $\left(\nabla f_{qk,1}(\mathbf{x}^{\star}) + \mu_{g_{qk}}(\mathbf{x}^{\star})\right)^T (\mathbf{x} - \mathbf{x}^{\star})$ of (4.63).

(c) If $p_q^{\star}(k) = 0$ and $M_{qk}(\mathbf{x}^{\star}) < 0$, following the same argument of the previous case, from (4.59) and (4.60) it is clear that: for all $\mathbf{x} \in \mathcal{P}$

$$\left(\nabla f_{qk,1}(\mathbf{x}^{\star}) + \nabla f_{qk,2}(\mathbf{x}^{\star})\right)^T (\mathbf{x} - \mathbf{x}^{\star}) \leq 0.$$

Again, from (4.62) it follows that: for all $\mathbf{x} \in \mathcal{P}$

$$r_{qk}^{s\prime}(\mathbf{x}^{\star}; \mathbf{x} - \mathbf{x}^{\star}) = 0.$$

Note that, if $\lambda_{qk}^{\star} = 0$ then $\mu_{g_{qk}}(\mathbf{x}^{\star}) = -\nabla f_{qk,1}(\mathbf{x}^{\star})$, and from the equation above we obtain the desired correspondence between the value of the (q,k)-th directional derivative and that of the term $\left(\nabla f_{qk,1}(\mathbf{x}^{\star}) + \mu_{g_{qk}}(\mathbf{x}^{\star})\right)^T (\mathbf{x} - \mathbf{x}^{\star})$ of (4.63).

(d) Finally, if $p_q^\star(k) > 0$ and $M_{qk}(\mathbf{x}^\star) = 0$, unfortunately not much can be said. After simplification, it follows from (4.59) and (4.60) that

$$(\nabla f_{qk,1}(\mathbf{x}^\star) + \nabla f_{qk,2}(\mathbf{x}^\star))^T (\mathbf{x} - \mathbf{x}^\star) =$$
$$\sum_{r \neq q} \left(\frac{(H_{qq}(k)H_{re}(k) - H_{qe}(k)H_{rq}(k))p_q^\star(k)}{I_{qq}(k)(H_{qe}(k)p_q(k) + I_{qe}(k))} \right) (p_r(k) - p_r^\star(k))$$
$$+ \sum_{j=1}^{J} \left(\frac{(H_{qq}(k)\widehat{H}_{je}(k) - H_{qe}(k)\widehat{H}_{jq}(k))p_q^\star(k)}{I_{qq}(k)(H_{qe}(k)p_q(k) + I_{qe}(k))} \right) (\widehat{p}_j(k) - \widehat{p}_j^\star(k))$$

where

$$I_{qq}(k) \triangleq \sigma_q^2(k) + \sum_{r \neq q} H_{rq}(k)\, p_r^\star(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j^\star(k)$$
$$I_{qe}(k) \triangleq \sigma_q^2(k) + \sum_{r \neq q} H_{re}(k)\, p_r^\star(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j^\star(k).$$

As a result, no sharp conclusion can be established regarding $r_{qk}^{s\,\prime}(\mathbf{x}^\star; \mathbf{x} - \mathbf{x}^\star)$. Nevertheless, notice that in this case, without loss of optimality, we can set $p_q^\star(k) = 0$ whenever no other main user is allocating power on subchannel k since the secrecy sum-rate is not affected. For example, in a system with orthogonal subchannels (see, Subsection 4.5.2) setting $p_q^\star(k) = 0$ when $M_{qk}(\mathbf{x}^\star) = 0$ does not affect the secrecy rate of the corresponding user-subchannel pair (q, k).

The observations above lead to the following result establishing some connections between the critical points of the DC problem (4.54) (obtained from Algorithm 4.4) and the d-stationary solutions of the SISO secrecy sum-rate maximization problem (4.47). The proof of the next proposition is a direct consequence of the cases (I) - (III) outlined above.

Proposition 4.8. Suppose that $\mathbf{x}^\star$ is a critical point of the DC program (4.54),

(a) If $f_{qk,1}(\mathbf{x}^\star) + f_{qk,2}(\mathbf{x}^\star) \neq 0$ for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$, then

$\mathbf{x}^\star$ is a d-stationary solution of the SISO secrecy sum-rate maximization problem (4.47).

(b) If $\mathbf{x}^\star$ is such that for every pair (q,k) satisfying $f_{qk,1}(\mathbf{x}^\star) + f_{qk,2}(\mathbf{x}^\star) = 0$ the following conditions hold:

(i) $p_q^\star(k) = 0$ whenever $M_{qk}(\mathbf{x}^\star) = 0$; and,

(ii) the corresponding (q,k)-th subgradient $\mu_{g_{qk}}(\mathbf{x}^\star) \in \partial g_{qk}(\mathbf{x}^\star)$ satisfies

$$\mu_{g_{qk}}(\mathbf{x}^\star) = \begin{cases} -\nabla f_{qk,1}(\mathbf{x}^\star) & \text{if } p_q^\star(k) = 0 \text{ and } M_{qk}(\mathbf{x}^\star) \leq 0 \\ \nabla f_{qk,2}(\mathbf{x}^\star) & \text{if } p_q^\star(k) = 0 \text{ and } M_{qk}(\mathbf{x}^\star) \geq 0, \end{cases}$$

then $\mathbf{x}^\star$ is a d-stationary solution of the SISO secrecy sum-rate maximization problem (4.47).

4.5.1.4 Joint Optimization Approach

In the following discussion, we apply the joint optimization approach introduced in Section 4.3 to the SISO secrecy sum-rate maximization problem. We start by rewriting such a problem [cf. (4.51)] as a (smooth) joint optimization problem in the variables $(\mathbf{p}, \widehat{\mathbf{p}})$ and $\boldsymbol{\lambda} \triangleq \left((\lambda_{qk})_{q=1}^{Q}\right)_{k=1}^{N}$, that is

$$\underset{(\mathbf{p},\widehat{\mathbf{p}})\in\mathcal{P},\ \lambda_{qk}\in[0,1]\,\forall q,k}{\text{maximize}} \quad \sum_{q=1}^{Q}\sum_{k=1}^{N} \left(\lambda_{qk}\, f_{qk,1}(\mathbf{p},\widehat{\mathbf{p}}) + \lambda_{qk}\, f_{qk,2}(\mathbf{p},\widehat{\mathbf{p}})\right). \tag{4.65}$$

Let us customize Algorithm 4.3 to the program (4.65). From the approximating function introduced in equation (4.34), we define: given $\mathbf{x}^\nu \triangleq (\mathbf{p}^\nu, \widehat{\mathbf{p}}^\nu) \in \mathcal{P}$ and $\lambda_{qk}^\nu \in [0,1]$

$$\begin{aligned} \widetilde{r}_{qk}^{\,s,J}(\mathbf{x}, \lambda_{qk} : \mathbf{x}^\nu, \lambda_{qk}^\nu) \triangleq{}& f_{qk,1}(\mathbf{x}) + \lambda_{qk}(f_{qk,1}(\mathbf{x}^\nu) + f_{qk,2}(\mathbf{x}^\nu)) \\ &+ \lambda_{qk}^\nu \nabla f_{qk,2}(\mathbf{x}^\nu)^T\mathbf{x} + (\lambda_{qk}^\nu - 1)\nabla f_{qk,1}(\mathbf{x}^\nu)^T\mathbf{x}, \end{aligned}$$

for all $q = 1,\ldots,Q$ and $k = 1,\ldots,N$. Then, from (4.35) the JO-based algorithm consists in solving iteratively the following sequence of strongly concave problems: given the vectors $\mathbf{x}^\nu \in \mathcal{P}$ and $\boldsymbol{\lambda}^\nu \triangleq \left((\lambda_{qk}^\nu)_{q=1}^{Q}\right)_{k=1}^{N}$ with each $\lambda_{qk}^{\nu} \in [0,1]$

$$\begin{aligned}\left(\mathbf{x}^{\nu+1}, \boldsymbol{\lambda}^{\nu+1}\right) \triangleq \underset{\mathbf{x}\in\mathcal{P},\, \lambda_{qk}\in[0,1]\,\forall q,k}{\operatorname{argmax}} \quad & \sum_{q=1}^{Q}\sum_{k=1}^{N} \widetilde{r}_{qk}^{\,s,J}(\mathbf{x}, \lambda_{qk} : \mathbf{x}^{\nu}, \lambda_{qk}^{\nu}) \\ & - \frac{\tau}{2}\|(\mathbf{x}, \boldsymbol{\lambda}) - (\mathbf{x}^{\nu}, \boldsymbol{\lambda}^{\nu})\|^2.\end{aligned} \tag{4.66}$$

In Algorithm 4.5 we introduce formally this approach. Notice that, in Step 2 of this iterative scheme we exploited the separability of the problem (4.66) in the variables $\mathbf{x}$ and $\boldsymbol{\lambda}$, as already pointed out in Section 4.4 for the general case. As a result, the update of the variable $\boldsymbol{\lambda}$ in (4.67) requires the solution of a scalar quadratic maximization problem for each $q = 1, \ldots, Q$ and $k = 1, \ldots, N$, whose solution can be found in closed form. Similarly, we recall that, for the DC-based algorithm, the update of this variable requires the solution of a scalar linear program, see equation (4.55).

It is easy to check that the requirements in part (a) of Proposition 4.6 are readily satisfied by our application, and if τ is chosen accordingly to condition (b) of the aforementioned proposition, then Algorithm 4.5 converges to stationary solutions of (4.65).

Algorithm 4.5: JO-based Algorithm for the SISO Secrecy Sum-Rate Maximization Problem

---

Data: $\tau > 0$, $\mathbf{x}^0 \triangleq (\mathbf{p}^0, \widehat{\mathbf{p}}^0) \in \mathcal{P}$, and $\lambda_{qk}^0 \in [0,1]$ for all $q = 1, \ldots, Q$ and $k = 1, \ldots, N$. Set $\nu = 0$.

(`S.1`): If $(\mathbf{x}^{\nu}, \boldsymbol{\lambda}^{\nu})$ satisfies a termination criterion, STOP.

(`S.2`): Compute: for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$

$$\lambda_{qk}^{\nu+1} \triangleq \underset{\lambda_{qk}\in[0,1]}{\operatorname{argmax}} \left[\lambda_{qk}\left(f_{qk,1}(\mathbf{x}^{\nu}) + f_{qk,2}(\mathbf{x}^{\nu})\right) - \frac{\tau}{2}\|\lambda_{qk} - \lambda_{qk}^{\nu}\|^2\right], \tag{4.67}$$

$$\begin{aligned}\mathbf{x}^{\nu+1} \triangleq \underset{\mathbf{x}\in\mathcal{P}}{\operatorname{argmax}} \sum_{q=1}^{Q}\sum_{k=1}^{N} & \left[\, f_{qk,1}(\mathbf{x}) + \lambda_{qk}^{\nu}\nabla f_{qk,2}(\mathbf{x}^{\nu})^T\mathbf{x} \right. \\ & \left. + (\lambda_{qk}^{\nu} - 1)\nabla f_{qk,1}(\mathbf{x}^{\nu})^T\mathbf{x}\,\right] - \frac{\tau}{2}\|\mathbf{x} - \mathbf{x}^{\nu}\|^2.\end{aligned} \tag{4.68}$$

(`S.3`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

We conclude this section by briefly discussing the relation between a stationary solution of the maximization problem (4.65) and a d-stationary solution of (4.47). This relation is summarized in Proposition 4.9, whose proof is very similar to that of its counterpart in Proposition 4.8, thus it is omitted.

Proposition 4.9. Suppose that $\left(\mathbf{x}^\star,\, \boldsymbol{\lambda}^\star \triangleq \left((\lambda^\star_{qk})_{q=1}^{Q}\right)_{k=1}^{N}\right)$ is a stationary solution of the joint optimization problem (4.65),

(a) If $f_{qk,1}(\mathbf{x}^\star)+f_{qk,2}(\mathbf{x}^\star) \neq 0$ for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$, then $\mathbf{x}^\star$ is a d-stationary solution of the SISO secrecy sum-rate maximization problem (4.47).

(b) If $\mathbf{x}^\star$ is such that for every pair (q, k) satisfying $f_{qk,1}(\mathbf{x}^\star) + f_{qk,2}(\mathbf{x}^\star) = 0$ the following conditions hold:

   (i) $p^\star_q(k) = 0$ whenever $M_{qk}(\mathbf{x}^\star) = 0$; and,

   (ii) $\lambda^\star_{qk}$ satisfies

$$\lambda^\star_{qk} = \begin{cases} 0 & \text{if } p^\star_q(k) = 0 \text{ and } M_{qk}(\mathbf{x}^\star) \leq 0 \\ 1 & \text{if } p^\star_q(k) = 0 \text{ and } M_{qk}(\mathbf{x}^\star) \geq 0, \end{cases}$$

then $\mathbf{x}^\star$ is a d-stationary solution of the SISO secrecy sum-rate maximization problem (4.47).

4.5.2 Multi-orthogonal Subchannels Secrecy Rate Game

This section presents an extension to the secrecy rate game considered in Chapter 3 - Subsection 3.5.1. The system model studied in the cited reference considers the scenario where each legitimate user communicates over a *single* orthogonal subchannel. In the forthcoming discussion, we consider the more general setting where the legitimate users communicate along *multiple* orthogonal subchannels.

4.5.2.1 System Model and Notation

Unless stated the contrary, throughout this section, we follow the same notation of Subsection 4.5.1. Similarly, we consider the same wireless communication system, but here we assume OFDMA transmissions for the legitimate users. More precisely, we introduce the following *orthogonal subchannels assumption:* each legitimate transmitter-receiver pair q communicates over a set of assigned subchannels denoted by $\mathcal{K}_q$. For every $q = 1, \ldots, Q$, the subchannels assignment sets $\mathcal{K}_q$ satisfy the properties:

(i) $\mathcal{K}_q \neq \emptyset$ for all $q = 1, \ldots, Q$,

(ii) $\mathcal{K}_q \cap \mathcal{K}_r = \emptyset$ for all $r \neq q$, and

(iii) $\bigcup_{q'=1}^{Q} \mathcal{K}_{q'} = \{1, \ldots, N\}$.

It is worth mentioning that orthogonal transmissions in the form of OFDMA is the most used standard in current and future wireless communications systems; examples are: WiMaX (IEEE 802.16), MBWA (IEEE 802.20), WRAN (IEEE 802.22), and 4G (LTE). Hence, a deep study of the power allocation problem in the OFDMA setting is of great importance.

Under the OFDMA assumption and letting $\mathbf{p}_q \triangleq (p_q(k))_{k\in\mathcal{K}_q}$ and $\widehat{\mathbf{p}}_q \triangleq \left(\widehat{\mathbf{p}}_q(k) \triangleq (\widehat{p}_j(k))_{j=1}^{J}\right)_{k\in\mathcal{K}_q}$, the secrecy rate of user q is given by

$$
\begin{aligned}
r_q^s\left(\mathbf{p}_q, \widehat{\mathbf{p}}_q\right) &\triangleq \sum_{k\in\mathcal{K}_q} r_{qk}^s(p_q(k), \widehat{\mathbf{p}}_q(k)) \\
&= \sum_{k\in\mathcal{K}_q} \left[r_{qqk}(p_q(k), \widehat{\mathbf{p}}_q(k)) - r_{qek}(p_q(k), \widehat{\mathbf{p}}_q(k))\right]^+
\end{aligned}
$$

where for every $k \in \mathcal{K}_q$

$$
\begin{aligned}
r_{qqk}(p_q(k), \widehat{\mathbf{p}}_q(k)) &\triangleq \log\left(1 + \frac{H_{qq}(k)\, p_q(k)}{\sigma_q^2(k) + \sum_{j=1}^{J} \widehat{H}_{jq}(k)\, \widehat{p}_j(k)}\right) \\
r_{qek}(p_q(k), \widehat{\mathbf{p}}_q(k)) &\triangleq \log\left(1 + \frac{H_{qe}(k)\, p_q(k)}{\sigma_q^2(k) + \sum_{j=1}^{J} \widehat{H}_{je}(k)\, \widehat{p}_j(k)}\right)
\end{aligned}
$$

Notice that in the equation above the interference terms of the legitimate users $r \neq q$ are no longer present in comparison with the scenario studied in the previous section [refer to Equation (4.46)].

4.5.2.2 Problem Formulation

The system design is formulated as a noncooperative game [76] where the main users are the players who aim to maximize their own secrecy rate with the help of the friendly jammers. Formally, each legitimate user q, anticipating $(\widehat{\mathbf{p}}_r)_{r\neq q}$ solves the following maximization problem:

$$\mathcal{G}: \quad \begin{array}{ll} \underset{(\mathbf{p}_q,\widehat{\mathbf{p}}_q)\geq \mathbf{0}}{\text{maximize}} & r_q^s\left(\mathbf{p}_q,\widehat{\mathbf{p}}_q\right) \\ \text{subject to:} & \left.\begin{array}{l} \displaystyle\sum_{k\in\mathcal{K}_q} p_q(k) \leq P_q^{\max}, \\ \displaystyle\sum_{k'=1}^{N} \widehat{p}_j(k') \leq \widehat{P}_j^{\max}, \ \forall\, j = 1,\ldots,J. \end{array}\right\} \triangleq \mathcal{P}_q(\widehat{\mathbf{p}}_{-q}) \end{array} \tag{4.69}$$

The game whose q-th optimization problem is given by (4.69) will be referred as the multi-orthogonal subchannels secrecy rate game $\mathcal{G}$. It is worth stressing some characteristics of $\mathcal{G}$ that make it challenging:

i) the players' objective functions are nondifferentiable and nonconcave; and,

ii) the feasible set $\mathcal{P}_q$ of the optimization problem above depends on $\widehat{\mathbf{p}}_{-q} \triangleq (\widehat{\mathbf{p}}_r)_{r\neq q}$, i.e. the friendly jammers' power allocation over the rest of legitimate users' subchannels.

This last feature of $\mathcal{G}$, classifies it as an instance of the well-known Generalized Nash Equilibrium Problems (GNEP) with shared constraints [27]. A solution of $\mathcal{G}$ is the Generalized Nash Equilibrium (GNE); the following definition formalizes this concept.

Definition 4.5. A strategy profile $\mathbf{x}^\star \triangleq \left(\mathbf{x}_q^\star \triangleq (\mathbf{p}_q^\star, \widehat{\mathbf{p}}_q^\star)\right)_{q=1}^{Q}$ is a GNE of the GNEP $\mathcal{G}$ if, for all $q = 1,\ldots,Q$, the following holds: $\mathbf{x}_q^\star \in \mathcal{P}_q(\widehat{\mathbf{p}}_{-q}^\star)$ and

$$r_q^s(\mathbf{x}_q^\star) \geq r_q^s(\mathbf{x}_q), \quad \forall\, \mathbf{x}_q \in \mathcal{P}_q(\widehat{\mathbf{p}}_{-q}^\star).$$

It is not difficult to check that the game $\mathcal{G}$ is of the potential type [31]. Let $\mathbf{p} \triangleq (\mathbf{p}_q)_{q=1}^{Q}$ and $\widehat{\mathbf{p}} \triangleq (\widehat{\mathbf{p}}_q)_{q=1}^{Q}$ then, the potential function of $\mathcal{G}$ is given by

$$r^s(\mathbf{p}, \widehat{\mathbf{p}}) \triangleq \sum_{q=1}^{Q} r_q^s\left(\mathbf{p}_q, \widehat{\mathbf{p}}_q\right).$$

Consider the associated social problem

$$(P): \quad \underset{(\mathbf{p},\widehat{\mathbf{p}}) \in \mathcal{P}^{\text{ort}}}{\text{maximize}} \; r^s(\mathbf{p}, \widehat{\mathbf{p}}), \tag{4.70}$$

where the feasible set is given by

$$\mathcal{P}^{\text{ort}} \triangleq \left\{ (\mathbf{p}, \widehat{\mathbf{p}}) \geq \mathbf{0} \,:\, \sum_{k \in \mathcal{K}_q} p_q(k) \leq P_q^{\max} \quad \forall\, q \text{ and, } \sum_{k'=1}^{N} \widehat{p}_j(k') \leq \widehat{P}_j^{\max} \quad \forall\, j \right\}.$$

For this type of games, it is well-known that any optimal solution of (P) is a GNE of $\mathcal{G}$. It is clear that (P) has a solution, therefore a GNE of $\mathcal{G}$ must exist. The following proposition summarizes this result.

Proposition 4.10. A GNE of the multi-orthogonal channels secrecy rate game $\mathcal{G}$ always exists.

Due to the intrinsic characteristics of $\mathcal{G}$, the computation of a GNE of this game is challenging. Thus, toward deriving a distributed algorithm for finding *practical solutions* of such a game, we focus on a relaxed equilibrium concept. Namely, we rely on the concept of B-Quasi GNE (B-QGNE) introduced in Chapter 3. Roughly speaking, a B-QGNE is a stationary solution of the GNEP based on the B(ouligand)-derivative [28]. The next definition gives a formal description of this concept.

Definition 4.6. A strategy profile $\mathbf{x}^\star \triangleq \left(\mathbf{x}_q^\star \triangleq (\mathbf{p}_q^\star, \widehat{\mathbf{p}}_q^\star)\right)_{q=1}^{Q}$ is a B-QGNE of the GNEP $\mathcal{G}$ if, for all $q = 1, \ldots, Q$, the following holds: $\mathbf{x}_q^\star \in \mathcal{P}_q(\widehat{\mathbf{p}}_{-q}^\star)$ and

$$r_q^{s\prime}(\mathbf{x}_q^\star; \mathbf{x}_q - \mathbf{x}_q^\star) \leq 0, \quad \forall\, \mathbf{x}_q \in \mathcal{P}_q(\widehat{\mathbf{p}}_{-q}^\star).$$

In the remaining discussion, we turn our attention to the distributed computation of a B-QGNE of $\mathcal{G}$. In pursuance of this goal, we capitalize on the potential structure of $\mathcal{G}$. The next proposition establishes the existence of a B-QGNE of $\mathcal{G}$ and summarizes some connections between the game $\mathcal{G}$ and the multiplayer maximization problem (P) defined in (4.70).

Proposition 4.11. Given the multi-orthogonal channels secrecy rate game $\mathcal{G}$ [cf. (4.69)] and the social problem (P) [cf. (4.70)], the following hold:

(a) A B-QGNE of the multi-orthogonal channels secrecy rate game $\mathcal{G}$ always exists.

(b) If $\mathbf{x}^\star$ is an optimal solution of (P), then $\mathbf{x}^\star$ is a GNE of $\mathcal{G}$.

(c) If $\mathbf{x}^\star$ is a d-stationary solution of (P), then $\mathbf{x}^\star$ is a B-QGNE of $\mathcal{G}$.

(d) If $\mathbf{x}^\star$ is a B-QGNE of $\mathcal{G}$ and there exists common multipliers of the shared constraints $\sum_{k=1}^{N} \widehat{p}_j(k) \leq \widehat{P}_j^{\max}$ $j = 1, \ldots, J$ for all players, then $\mathbf{x}^\star$ is a d-stationary solution of (P).

Proof. (a) Since the players' optimization problems of the game $\mathcal{G}$ have feasible sets of the polyhedral type, every GNE of this game is a B-QGNE. From Proposition 4.10, a GNE of $\mathcal{G}$ always exists, thus the existence of a B-QGNE follows readily.

(b) This follows immediately since $\mathcal{G}$ is an exact potential game.

(c) The proof of this statement is similar to that of Proposition 3.6(c) in Chapter 3, and thus we omit further details.

(d) This follows readily under the common multipliers assumption. □

Proposition 4.11(c) along with the observation that the social problem (P) is an instance of the MSM problem (4.2) pave the way on deriving iterative algorithms for attempting the computation of a B-QGNE of $\mathcal{G}$. Either Algorithm 4.4 or 4.5 can be applied toward achieving that goal. In what follows, we focus on the DC-Programming based approach, nevertheless a similar analysis can be done for the JO-based scheme. Algorithm 4.4 is shown to be convergent to *critical points* of the DC-reformulation of (P) given by

$$(P_{DC}): \underset{(\mathbf{p},\widehat{\mathbf{p}})\in\mathcal{P}^{\text{ort}}}{\text{maximize}} \sum_{q=1}^{Q} \sum_{k\in\mathcal{K}_q} f_{qk,1}(p_q(k), \widehat{\mathbf{p}}_q(k)) + g_{qk}(p_q(k), \widehat{\mathbf{p}}_q(k)), \quad (4.71)$$

where $f_{qk,1}$ and g_{qk} are defined similarly as those in (4.50) and (4.53) respectively, by imposing the orthogonality assumption. As a side note, in what follows we omit these sort of definitions since they are easily obtained from those in Subsection 4.5.1 by setting to zero the legitimate users' interference terms. We emphasize that the relation between the problems (P) and (P_{DC})

has been deeply explored in previous sections. It is clear that, (P) and (P_{DC}) are equivalent in terms of globally optimal and d-stationary solutions, hence from Proposition 4.11(b) any optimal solution of (P_{DC}) is GNE of the game $\mathcal{G}$. Additionally, a critical point of (P_{DC}) satisfying either conditions (a) or (b) of Proposition 4.8 is a d-stationary solution of (P), thus from Proposition 4.11(c) this point corresponds to a B-QGNE of $\mathcal{G}$. We remark that, in practice, such conditions are easily satisfied. As a result, practically speaking, we can apply Algorithm 4.4 to compute a B-QGNE of the multi-orthogonal channels secrecy game $\mathcal{G}$. For the sake of clarity, Figure 4.4 summarizes the connections between the game $\mathcal{G}$ and the multiuser optimization problems (P) and (P_{DC}).

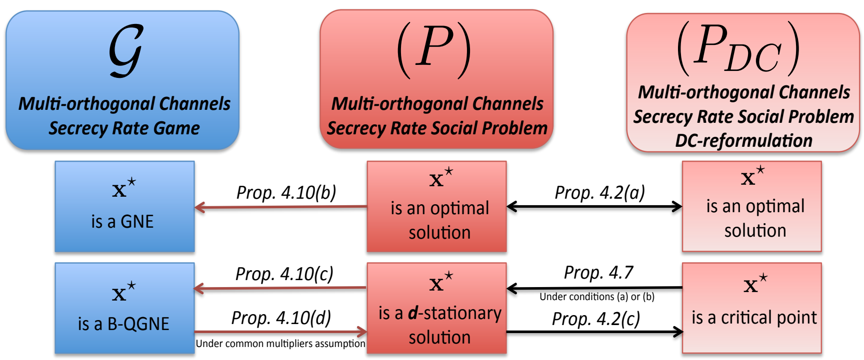

Figure 4.4: Connections between the multi-orthogonal secrecy rate game $\mathcal{G}$, the social problem (P) and its DC-reformulation (P_{DC}).

4.5.2.3 Algorithmic Design

Algorithm 4.4 is *centralized*, that is, its implementation requires coordination among the different entities in the network. An important implication of the OFDMA assumption is that the objective function of the optimization problem (4.70) is separable in the legitimate users' variables i.e. the secrecy rate of user q (r_q^s) depends on his private variables $\mathbf{x}_q \triangleq (\mathbf{p}_q, \widehat{\mathbf{p}}_q)$ alone. However, the presence of coupling constraints (i.e. the friendly jammers' power budget) complicates the design of distributed algorithms. Following the approach in Chapter 3, we cope with this issue by noticing that the subproblems in step

(S.4) of Algorithm 4.4 i.e. given $\mathbf{x}^{\nu} \triangleq \left(\mathbf{x}_q^{\nu} \triangleq \left(\mathbf{p}_q^{\nu}, \widehat{\mathbf{p}}_q^{\nu}\right)\right)_{q=1}^{Q} \in \mathcal{P}^{\mathtt{ort}}$

$$\begin{aligned}\mathbf{x}^{\nu+1} \triangleq \underset{\mathbf{x}\triangleq(\mathbf{x}_q)_{q=1}^{Q}\in\mathcal{P}^{\mathtt{ort}}}{\operatorname{argmax}} \quad & \sum_{q=1}^{Q}\sum_{k\in\mathcal{K}_q}[\, f_{qk,1}(p_q(k), \widehat{\mathbf{p}}_q(k)) \\ & + \mu_{g_{qk}}(p_q^{\nu}(k), \widehat{\mathbf{p}}_q^{\nu}(k))^T(p_q(k), \widehat{\mathbf{p}}_q(k)) - \frac{\tau_q}{2}\|\mathbf{x}_q - \mathbf{x}_q^{\nu}\|^2\,],\end{aligned} \tag{4.72}$$

can be solved in a fairly distributed way if the friendly jammers' power budget constraints are dualized. Let the set of private constraints be denoted by $\mathcal{X}_q \triangleq \left\{(\mathbf{p}_q, \widehat{\mathbf{p}}_q) \geq \mathbf{0} \,:\, \sum_{k\in\mathcal{K}_q} p_q(k) \leq P_q^{\max}\right\}$ for every $q = 1, \ldots, Q$, and let $\mathcal{X} \triangleq \Pi_{q=1}^{Q}\mathcal{X}_q$. Then, the dual problem associated with (4.72) i.e. given $\mathbf{x}^{\nu} \in \mathcal{P}^{\mathtt{ort}}$

$$\underset{\boldsymbol{\pi}\triangleq(\pi_j)_{j=1}^{J}\geq\mathbf{0}}{\operatorname{minimize}} \quad d\left(\boldsymbol{\pi}; \mathbf{x}^{\nu}\right), \tag{4.73}$$

where the dual function is defined as

$$\begin{aligned} d\left(\boldsymbol{\pi}; \mathbf{x}^{\nu}\right) \triangleq \underset{(\mathbf{p},\widehat{\mathbf{p}})\,\in\,\mathcal{X}}{\operatorname{maximum}} \quad & \sum_{q=1}^{Q}\sum_{k\in\mathcal{K}_q}(\, f_{qk,1}(p_q(k), \widehat{\mathbf{p}}_q(k)) \\ & + \mu_{g_{qk}}(p_q^{\nu}(k), \widehat{\mathbf{p}}_q^{\nu}(k))^T(p_q(k), \widehat{\mathbf{p}}_q(k))\,) \\ & - \frac{\tau_q}{2}\|\mathbf{x}_q - \mathbf{x}_q^{\nu}\|^2 - \sum_{j=1}^{J}\pi_j\left(\sum_{k=1}^{N}\widehat{p}_j(k) - \widehat{P}_j^{\max}\right),\end{aligned}$$

can be solved using, for example, gradient projection algorithms [11]. Notice that, the dual function d is differentiable on $\mathbb{R}_+^J$ (as a direct consequence of [11, Prop. 4.5.1]). From the cartesian structure of the set $\mathcal{X}$, the strongly concave optimization problem above is separable in the legitimate users' variables. Let its unique solution be denoted by $\widehat{\mathbf{x}}(\boldsymbol{\pi}; \mathbf{x}^{\nu}) \triangleq \left(\widehat{\mathbf{x}}_q(\boldsymbol{\pi}; \mathbf{x}_q^{\nu})\right)_{q=1}^{Q}$, where for every $q = 1, \ldots, Q$

$$\begin{aligned}\widehat{\mathbf{x}}_q(\boldsymbol{\pi}; \mathbf{x}_q^{\nu}) \triangleq \underset{\mathbf{x}_q\,\in\,\mathcal{X}_q}{\operatorname{argmax}} \quad & \sum_{k\in\mathcal{K}_q}[\, f_{qk,1}(p_q(k), \widehat{\mathbf{p}}_q(k)) \\ & + \mu_{g_{qk}}(p_q^{\nu}(k), \widehat{\mathbf{p}}_q^{\nu}(k))^T(p_q(k), \widehat{\mathbf{p}}_q(k))\,] \\ & - \frac{\tau_q}{2}\|\mathbf{x}_q - \mathbf{x}_q^{\nu}\|^2 - \sum_{j=1}^{J}\pi_j\sum_{k\in\mathcal{K}_q}\widehat{p}_j(k).\end{aligned} \tag{4.74}$$

Then, the proposed *distributed* algorithm is a double loop scheme with communication between them. In essence, the inner loop consists in solving problem (4.72) via a dual approach. An instance of the resulting scheme, based on gradient projection algorithms, is given in Algorithm 4.6. We remark that this dual approach makes possible the decentralization of the computations at the expense of introducing an additional loop into the algorithm.

The convergence of Algorithm 4.6 to critical points of the problem (P_{DC}) can be established as follows. Given any $\boldsymbol{\tau} \triangleq (\tau_q)_{q=1}^{Q} > 0$ and if the step-size sequence $\{\alpha^t\}$ is chosen accordingly to Theorem 3.4 in Chapter 3 then the sequence $\{\boldsymbol{\pi}^t\}$ converges to a solution of the dual problem (4.73). Since there is no duality gap (recall that, the feasible set $\mathcal{P}^{\texttt{ort}}$ is polyhedral), then the sequence $\{(\mathbf{p}^{\nu,t}, \widehat{\mathbf{p}}^{\nu,t})\}$ converges to the unique solution of (4.72). Therefore, the outer loop must converge to a critical point of (P_{DC}) in the sense of Proposition 4.3.

Algorithm 4.6 : Distributed DC-based Algorithm for the OFDMA SISO Secrecy Sum-Rate Maximization Problem

---

Data: $\boldsymbol{\tau} \triangleq (\tau_q)_{q=1}^{Q} > 0$, $\{\alpha^t\} > 0$ and $\mathbf{x}^0 \triangleq (\mathbf{p}^0, \widehat{\mathbf{p}}^0) \in \mathcal{P}$. Set $\nu = 0$.

(`S.1`): If $\mathbf{x}^\nu \triangleq (\mathbf{p}^\nu, \widehat{\mathbf{p}}^\nu)$ satisfies a termination criterion, STOP.

(`S.2`): The main users $q = 1, \ldots, Q$ compute in parallel $\lambda_{qk}^{\star,\nu}$ for every $k \in \mathcal{K}_q$ as in (4.55).

(`S.3`): The main users $q = 1, \ldots, Q$ compute in parallel $\mu_{g_{qk}}(p_q^\nu(k), \widehat{\mathbf{p}}_q^\nu(k))$ for every $k \in \mathcal{K}_q$ as in (4.56).

(`S.4a`): Choose any $\boldsymbol{\pi}^0 \triangleq (\pi_j^0)_{j=1}^{J} \geq \mathbf{0}$. Set $t = 0$.

(`S.4b`): If $\boldsymbol{\pi}^t \triangleq (\pi_j^t)_{j=1}^{J}$ is a solution of (4.73), set $\widehat{\mathbf{x}}(\mathbf{x}^\nu) = \left(\mathbf{p}_q^{\nu,t}, \widehat{\mathbf{p}}_q^{\nu,t}\right)_{q=1}^{Q}$, and go to (`S.5`).

(`S.4c`): The main users $q = 1, \ldots, Q$ compute in parallel $\left(\mathbf{p}_q^{\nu,t}, \widehat{\mathbf{p}}_q^{\nu,t}\right)$ by solving (4.74).

(`S.4d`): The friendly jammers $j = 1, \ldots, J$ update in parallel $\boldsymbol{\pi} \triangleq (\pi_j)_{j=1}^{J}$ by computing

$$\pi_j^{t+1} \triangleq \left[\pi_j^t + \alpha^t \left(\sum_{k=1}^{N} \widehat{p}_j^{\nu,t}(k) - \widehat{P}_j^{\max}\right)\right]^+. \tag{4.75}$$

(`S.5`): Set $\mathbf{x}^{\nu+1} = \widehat{\mathbf{x}}(\mathbf{x}^\nu)$.

(`S.6`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

Remark 4.4. *(On the Implementation of Algorithm 4.6).* Provided that the CSI is available at the main users's sides, Algorithm 4.6 can be implemented in a fairly distributed way as follows. At any inner loop iteration $t \geq 0$, given the multiplier $\boldsymbol{\pi}^t \triangleq (\pi_j^t)_{j=1}^J \geq \mathbf{0}$, all the legitimate users update *in parallel* their power allocation vectors $\left(\mathbf{p}_q^{\nu,t}, \widehat{\mathbf{p}}_q^{\nu,t}\right)$ by solving the strongly concave subproblem (4.74). Notice that, the computations of the subgradients $\mu_{g_{qk}}(p_q^{\nu}(k), \widehat{\mathbf{p}}_q^{\nu}(k))$ are *local* and thus, steps (`S.2`) and (`S.3`) do not require any signaling between the network users. After updating their power profiles, each legitimate user q communicates to the friendly jammers the fraction of power required to secure his transmission. Then, each friendly jammer j updates in parallel and independently the multiplier π_j^t via the inexpensive scalar projection in (4.75), where $\alpha^t > 0$ denotes a step size. Finally, the updated price vector $\boldsymbol{\pi}^{t+1}$ is broadcasted to the main users. This terminates one cycle of execution of the inner loop. After a successful termination of this loop, in the outer loop each main user updates his corresponding power allocation by replacing the current one with the inner loop's output.

To conclude the discussion, it is important to highlight that the theory developed to analyze the MSM problem leads to an analysis of the secrecy rate game for a more general system model than the one presented in Chapter 3. In particular, by following the MSM approach we were able to get rid of the *smooth game* formulation used in Chapter 3 to derive a distributed algorithm computing a (restricted) B-QGNE of the single channel case game. Nevertheless, the approach followed in this section relies on two aspects: (i) a second optimization problem, namely on the DC-reformulation of the optimization problem (P), rather than on another game formulation; and, (ii) the computation of a relaxed equilibrium point of the multi-orthogonal secrecy rate game depends on the critical points (produced by the DC-based Algorithm 4.6) satisfying either conditions (a) or (b) of Proposition 4.8. Finally, it is worth mentioning that the smooth game approach used in the cited chapter is suitable for the single channel case, but too restrictive for the multi-orthogonal subchannels setting.

4.5.3 Alternative System Designs

The SISO secrecy sum-rate maximization problem may lead to unfair power allocations with possibly some users not transmitting at all. There are certain communication systems where a minimum Quality of Service (QoS) needs to be guaranteed. Consequently, the aforementioned system design is not feasible in those particular situations. In this section, we briefly consider two different designs that aim to overcome these difficulties. First, we consider the well-known Max-Min fairness case, and second, we extend the design in (4.47) by introducing QoS constraints into this model. Interestingly, the optimization problems associated with these models can be cast into the general DC program (4.27), thus we can easily derive iterative algorithms converging to critical points of these problems.

4.5.3.1 Max-Min Fairness in SISO Secrecy Sum-Rate

Given a *secrecy rate profile* vector $\boldsymbol{\beta} \triangleq (\beta_q)_{q=1}^{Q}$ satisfying the conditions $\boldsymbol{\beta} > \mathbf{0}$ and $\sum_{q=1}^{Q} \beta_q = 1$, the Max-Min fairness system design is given by the following maximization problem:

$$\begin{aligned} &\underset{r,\,(\mathbf{p},\widehat{\mathbf{p}})\in\mathcal{P}}{\text{maximize}} \quad r \\ &\text{subject to:} \\ &\qquad \sum_{k=1}^{N} r_{qk}^{s}(\mathbf{p},\widehat{\mathbf{p}}) \geq \beta_q\, r \qquad \forall\, q = 1,\ldots,Q. \end{aligned} \tag{4.76}$$

Invoking the DC-decomposition of r_{qk}^{s} obtained in (4.52), it is clear that the constraints in the problem above can be easily rewritten as DC functions. Namely, the DC-reformulation of the Max-Min fairness problem is:

$$\begin{aligned} &\underset{r,\,(\mathbf{p},\widehat{\mathbf{p}})\in\mathcal{P}}{\text{maximize}} \quad r \\ &\text{subject to:} \\ &\qquad \sum_{k=1}^{N} f_{qk,1}(\mathbf{p},\widehat{\mathbf{p}}) - (-g_{qk}(\mathbf{p},\widehat{\mathbf{p}})) \geq \beta_q\, r \qquad \forall\, q = 1,\ldots,Q, \end{aligned} \tag{4.77}$$

where the functions $f_{qk,1}$ and g_{qk} are defined in (4.50) and (4.53), respectively. Clearly, the maximization problem above is an instance of (4.27), thus, Algorithm 4.2 can be used to compute critical points of this DC program. Algorithm 4.7 customizes Algorithm 4.2 to the Max-Min fairness problem;

where, following the approach described in Subsection 4.2.3, the nonconvex feasible set of (4.76), denoted by

$$\mathcal{P}^{\mathtt{mm}} \triangleq \left\{ r,\ (\mathbf{p}, \widehat{\mathbf{p}}) \in \mathcal{P}\ :\ \sum_{k=1}^{N} r_{qk}^{s}(\mathbf{p}, \widehat{\mathbf{p}}) \geq \beta_q\, r \quad \forall\, q = 1, \ldots, Q \right\},$$

is approximated at every $\mathbf{x}^{\nu} \triangleq (\mathbf{p}^{\nu}, \widehat{\mathbf{p}}^{\nu}) \in \mathcal{P}^{\mathtt{mm}}$ by the following convex set

$$\widetilde{\mathcal{P}}^{\mathtt{mm}}(\mathbf{x}^{\nu}) \triangleq$$
$$\left\{ r, (\mathbf{p}, \widehat{\mathbf{p}}) \in \mathcal{P} : -\left(\sum_{k=1}^{N} f_{qk,1}(\mathbf{x}) + g_{qk}(\mathbf{x}^{\nu}) + \mu_{g_{qk}}(\mathbf{x}^{\nu})^{T}(\mathbf{x} - \mathbf{x}^{\nu}) \right) + \beta_q r \leq 0,\ \forall q \right\}.$$

Note that, the strongly concave problem (4.78) in step (`S.4`) of Algorithm 4.7 has a simple objective function, namely, a scalar quadratic function. Besides, we emphasize that the convergence of Algorithm 4.7 to critical points of the maximization problem (4.77) is due to Proposition 4.4.

Algorithm 4.7 : DC-based Algorithm for the Max-Min Fairness in SISO Secrecy Sum-Rate

---

Data: $\tau > 0$ and $\left(r^{0}, \mathbf{x}^{0} \triangleq (\mathbf{p}^{0}, \widehat{\mathbf{p}}^{0}) \right) \in \mathcal{P}^{\mathtt{mm}}$. Set $\nu = 0$.

(`S.1`): If $\left(r^{\nu}, \mathbf{x}^{\nu} \triangleq (\mathbf{p}^{\nu}, \widehat{\mathbf{p}}^{\nu}) \right)$ satisfies a termination criterion, STOP.

(`S.2`): For $q = 1, \ldots, Q$ and $k = 1, \ldots, N$ compute $\lambda_{qk}^{\star,\nu}$ as in (4.55).

(`S.3`): For $q = 1, \ldots, Q$ and $k = 1, \ldots, N$ compute $\mu_{g_{qk}}(\mathbf{x}^{\nu})$ as in (4.56).

(`S.4`): Compute

$$(r^{\nu+1}, \mathbf{x}^{\nu+1}) \triangleq \underset{(r,\mathbf{x}) \in \widetilde{\mathcal{P}}^{\mathtt{mm}}(\mathbf{x}^{\nu})}{\operatorname{argmax}}\ r - \frac{\tau}{2}\,(r - r^{\nu})^{2}. \tag{4.78}$$

(`S.5`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

4.5.3.2 SISO Secrecy Sum-Rate Maximization with QoS constraints

An alternative solution to the possibly unfair solutions obtained from (4.47) is to incorporate QoS constraints into this model. In this system design, given the *secrecy rate profile* $\mathbf{s}^{\star} \triangleq (s_q^{\star})_{q=1}^{Q} \geq \mathbf{0}$ where $s_q^{\star}$ denotes the minimum secrecy rate required by user q, we consider the maximization problem:

$$\begin{aligned} &\underset{(\mathbf{p},\widehat{\mathbf{p}})\in\mathcal{P}}{\text{maximize}} \quad \sum_{q=1}^{Q}\sum_{k=1}^{N} r_{qk}^{s}(\mathbf{p},\widehat{\mathbf{p}}) \\ &\text{subject to:} \\ &\qquad \sum_{k=1}^{N} r_{qk}^{s}(\mathbf{p},\widehat{\mathbf{p}}) \geq s_q^{\star} \qquad \forall\, q=1,\ldots,Q. \end{aligned} \tag{4.79}$$

It is worth stressing that the secrecy rate profile is such that $s_q^{\star} \geq 0$ for all $q = 1,\ldots,Q$, thus we left open the possibility of having $s_q^{\star} = 0$ for some q. The latter case implies that the corresponding user q does not have a minimum QoS requirement. This is important in those communication systems where some legitimate users do not demand a minimum secrecy rate while others do, and, in the case where, due to the physical characteristics of the network, the secrecy rate of some users cannot be greater than zero.

Once again, from the nontrivial DC-decomposition of r_{qk}^{s} obtained in (4.52), it follows that the multiuser optimization problem (4.79) can be reformulated as a general DC program. This DC-reformulation is given by:

$$\begin{aligned} &\underset{(\mathbf{p},\widehat{\mathbf{p}})\in\mathcal{P}}{\text{maximize}} \quad \sum_{q=1}^{Q}\sum_{k=1}^{N} f_{qk,1}(\mathbf{p},\widehat{\mathbf{p}}) - (-g_{qk}(\mathbf{p},\widehat{\mathbf{p}})) \\ &\text{subject to:} \\ &\qquad \sum_{k=1}^{N} f_{qk,1}(\mathbf{p},\widehat{\mathbf{p}}) - (-g_{qk}(\mathbf{p},\widehat{\mathbf{p}})) \geq s_q^{\star} \quad \forall\, q=1,\ldots,Q, \end{aligned} \tag{4.80}$$

where the functions $f_{qk,1}$ and g_{qk} are defined in (4.50) and (4.53), respectively. Clearly, (4.80) is an instance of the optimization problem (4.27). Consequently, Algorithm 4.2 can be used to compute critical points of the above DC program. Algorithm 4.8 customizes Algorithm 4.2 to the secrecy sum-rate maximization problem with QoS constraints. Basically, the only variant from Algorithm 4.4 is in step (`S.4`). In this case, the nonconvex set of constraints in (4.79), denoted by

$$\mathcal{P}^{\mathsf{QoS}} \triangleq \left\{ (\mathbf{p},\widehat{\mathbf{p}}) \in \mathcal{P} \,:\, \sum_{k=1}^{N} r_{qk}^{s}(\mathbf{p},\widehat{\mathbf{p}}) \geq s_q^{\star} \quad \forall\, q=1,\ldots,Q \right\},$$

is approximated at every $\mathbf{x}^\nu \triangleq (\mathbf{p}^\nu, \widehat{\mathbf{p}}^\nu) \in \mathcal{P}^{\texttt{QoS}}$ by the convex set

$$\widetilde{\mathcal{P}}^{\texttt{QoS}}(\mathbf{x}^\nu) \triangleq$$
$$\left\{(\mathbf{p}, \widehat{\mathbf{p}}) \in \mathcal{P} : -\left(\sum_{k=1}^{N} f_{qk,1}(\mathbf{x}) + g_{qk}(\mathbf{x}^\nu) + \mu_{g_{qk}}(\mathbf{x}^\nu)^T(\mathbf{x} - \mathbf{x}^\nu)\right) + s_q^\star \leq 0,\ \forall q\right\}.$$

This gives rise to the strongly concave problem in (4.81), where both the objective function and the constraints are nonlinear, as opposed to the linearly constrained optimization problem in Step (`S.4`) of Algorithm 4.4. Finally, it is worth mentioning that the convergence of Algorithm 4.8 to critical points of (4.80) can be established by Proposition 4.4.

Algorithm 4.8: DC-based Algorithm for the SISO Secrecy Sum-Rate Maximization Problem with QoS Constraints

Data: $\tau > 0$ and $\mathbf{x}^0 \triangleq (\mathbf{p}^0, \widehat{\mathbf{p}}^0) \in \mathcal{P}^{\texttt{QoS}}$. Set $\nu = 0$.

(`S.1`): If $\mathbf{x}^\nu \triangleq (\mathbf{p}^\nu, \widehat{\mathbf{p}}^\nu)$ satisfies a termination criterion, STOP.

(`S.2`): For $q = 1, \ldots, Q$ and $k = 1, \ldots, N$ compute $\lambda_{qk}^{\star;\nu}$ as in (4.55).

(`S.3`): For $q = 1, \ldots, Q$ and $k = 1, \ldots, N$ compute $\mu_{g_{qk}}(\mathbf{x}^\nu)$ as in (4.56).

(`S.4`): Compute

$$\mathbf{x}^{\nu+1} \triangleq \underset{\mathbf{x} \in \widetilde{\mathcal{P}}^{\texttt{QoS}}(\mathbf{x}^\nu)}{\operatorname{argmax}} \sum_{q=1}^{Q} \sum_{k=1}^{N} \left(f_{qk,1}(\mathbf{x}) + \mu_{g_{qk}}(\mathbf{x}^\nu)^T \mathbf{x}\right) - \frac{\tau}{2}\|\mathbf{x} - \mathbf{x}^\nu\|^2. \tag{4.81}$$

(`S.5`): $\nu \leftarrow \nu + 1$ and go to (`S.1`).

4.5.4 Numerical Results

In this section, we present some numerical experiments in order to study the performance of the algorithms developed above for the power allocation problem in a physical layer based security model for the SISO system. More specifically, we compare these algorithms in terms of the secrecy sum-rate attained and the number of iterations required to achieved it. We also contrast our algorithms with other applicable schemes existing in the literature adapted to our formulations. At the same time, we analyze the impact that the introduction of friendly jammers has into the communication system's performance, as well as the secrecy sum-rate gain obtained by adopting the proposed schemes versus fixed power allocation policies. We start the analysis by describing the system setup and the algorithms' parameters used in all the experiments presented below.

System Setup. Unless stated the contrary, all the experiments were obtained under the following system settings. We assumed that all the legitimate users and friendly jammers are endowed with the same power budget i.e. $P_q^{\max} = \widehat{P}_j^{\max} = P$ for all $q = 1, \ldots, Q$ and all $j = 1, \ldots, J$. Similarly, the noise variances are such that $\sigma_q^2(k) = \sigma^2$ for all $q = 1, \ldots, Q$ and all $k = 1, \ldots, N$. We set the signal-to-noise ratio $\texttt{snr} \triangleq P/\sigma^2 = 3$ dB. Additionally, the positions of the main users, friendly jammers and eavesdropper were randomly generated within a square of unit area. The main users' channels were simulated as Finite Impulse Response (FIR) filters of order $L = 10$, whose taps are independent and identically distributed zero mean complex Gaussian random variables with variance $1/(d_{rq}^{\gamma}(L+1)^2)$, where d_{rq} denotes the distance between the transmitter of user r and the receiver of user q, and $\gamma = 1.8$ is the path loss exponent. The same approach was used to simulate the rest of channels in the system.

Recall that, in the OFDMA case, a set of subchannels $\mathcal{K}_q$ is assigned to every user $q = 1, \ldots, Q$ (refer to Subsection 4.5.2). In the results below dealing with this particular case, we generate 10 random subchannel assignments for each set of channel realizations, and then, we took the corresponding average over them. The cardinal of every set $\mathcal{K}_q$ is equal to Q/N whenever the remainder ρ after this division is equal to zero, otherwise ρ users will be assigned $\lfloor Q/N \rfloor + 1$ subchannels, while the rest $Q - \rho$ users will have sets with cardinal $\lfloor Q/N \rfloor$.

Algorithms Setup. For Algorithm 4.4 we set the regularization constant $\tau = 1/2$, and for Algorithm 4.6 we used $\tau_q = 1/2$ for all $q = 1, \ldots, Q$. All the algorithms were initialized using a uniform power allocation vector for both the legitimate users and the friendly jammers, and their execution is terminated when the absolute value of the difference of their corresponding objective values in two consecutive iterations becomes smaller than $1e-5$. For Algorithm 4.6, which is a double loop scheme, the inner loop is terminated when the difference of the norm of the price vectors in two consecutive rounds is less than $1e-4$.

Example 4.1. DC-Programming versus Joint Optimization Aproach. In Figure 4.5 we compare the two approaches proposed in this chapter to address the SISO secrecy sum-rate maximization problem under the general setting in (4.47) (blue line curves), and under the OFDMA model in (4.70) (red line curves). In these experiments, we fixed the number of subchannels to $N = 16$ and increase the number Q of legitimate users, while there are $J = Q/2$ friendly jammers. This comparison is done in terms of average secrecy sum-rate achieved and average number of iterations required to attain it. For the double loop algorithms, the average total number of iterations (including inner and outer loop iterations) is reported. Such averages were calculated over 50 independent channel realizations. The plots indicate that the average secrecy sum-rate achieved by the joint optimization approach (Algorithm 4.5, and its variant for the OFDMA case) is very similar to the one obtained from the DC-Programming based schemes (Algorithm 4.4, and Algorithm 4.6 for the OFDMA case); see Figure 4.5(a). Moreover, both approaches require (on average) almost the same number of iterations to converge; refer to Figure 4.5(b). As a result, these numerical experiments suggest that no significant performance difference (both in terms of achieved objective value and number of iterations) is observed between the DC and JO based schemes, at least for the resource allocation problem under consideration. This conclusion is in accordance with the observations made in Section 4.4.

Example 4.2. Secrecy Sum-Rate Gain. In Figure 4.6, we plot the average secrecy sum-rate (taken over 50 independent channel realizations) achieved by Algorithm 4.4 for the SISO secrecy sum-rate maximization problem [cf. (4.47)] (blue line curves), and by Algorithm 4.6 for the orthogonal subchannels case [cf. (4.70)] (red line curves), under the following three different scenarios:

(a) the number of subchannels is fixed to $N = 16$ while the number of main users and friendly jammers increases; when present there are $J = Q/2$ jammers (see, Figure 4.6(a));

(b) the number of main users is fixed to $Q = 4$ and when present there are $J = 2$ friendly jammers, while the number of subchannels is increased (see, Figure 4.6(b));

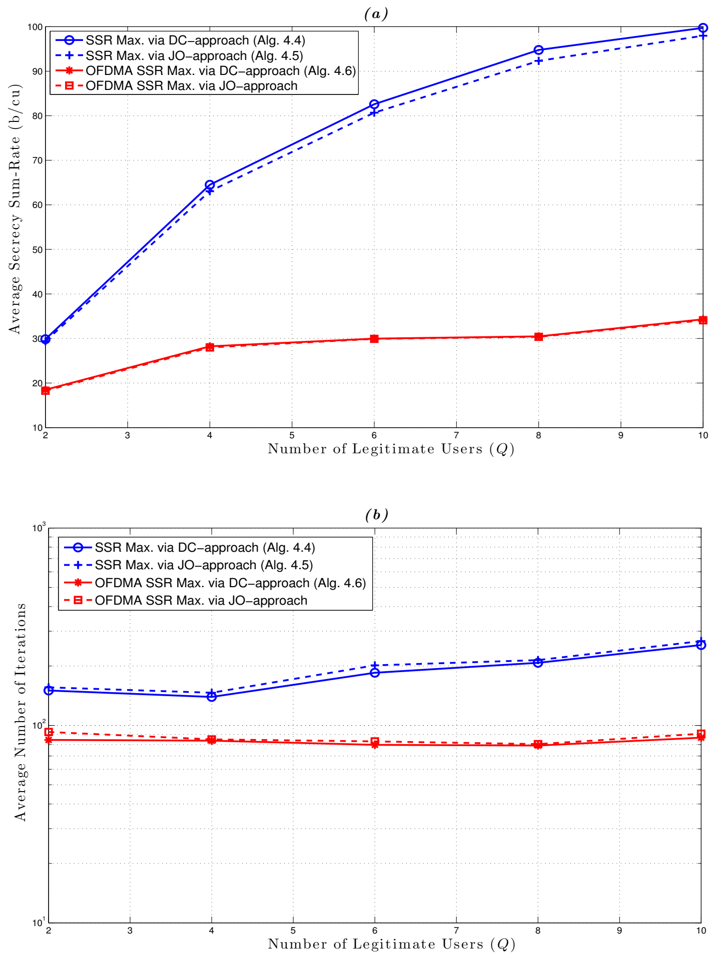

Figure 4.5: Comparison between the DC-Programming and the Joint Optimization (JO) based algorithms in terms of (a) average secrecy sum-rate (SSR), and (b) average number of iterations.

(c) the `snr` is increased, and the number of legitimate users, friendly jammers and subchannels are fixed to $Q = 4$, $J = 2$ and $N = 16$, respectively (see, Figure 4.6(c)).

For each of these tree scenarios, Figure 4.6 also shows the case $J = 0$ i.e. where no friendly jammers are present in the communication system (dashed line curves), and the simple situation where the uniform power policy is used to allocate the resources in the system (dotted line curves).

From these plots, it is clear that Algorithm 4.4 yields much higher secrecy sum-rates than those achievable by Algorithm 4.6 for the orthogonal subchannels case. The aforementioned behavior was expected since Algorithm 4.6 only optimizes the power allocation while the subchannels are randomly assigned, nevertheless the main advantage of this scheme is that it is distributed. Furthermore, the secrecy sum-rates obtained by the proposed system designs are significantly better than those coming from the fixed uniform power allocation policy. These plots also show that the introduction of friendly jammers into the communication system is beneficial for its performance, since the secrecy sum-rates obtained are higher than those attained for the $J = 0$ case. Notice that this gain becomes more significant for the case of orthogonal subchannels than for the general model; this is because, in the latter case, the legitimate users may also act as friendly jammers. As expected, this secrecy sum-rate gain is also more significant when the `snr` is increased (see, Figure 4.6(c)).

Example 4.3. Convergence Speed. Figure 4.7 shows the average number of iterations (taken over 50 independent channel realizations) required by Algorithm 4.4 and 4.6 to converge for the scenarios (a) and (b) described in Example 2. For Algorithm 4.6, a double loop scheme, we report the average number of iterations including both the inner and outer loop. Clearly, Algorithm 4.6 outperforms Algorithm 4.4 in terms of number of iteration, even more, the former algorithm is also distributed; however this improved performance comes at the expense of significantly smaller secrecy sum-rates achieved than those obtained from the centralized approach (see, Figure 4.6).

Example 4.4. Comparison Between our Algorithms 4.4, 4.6 and Other Available Schemes. Algorithms 4.4 and 4.6 are provable con-

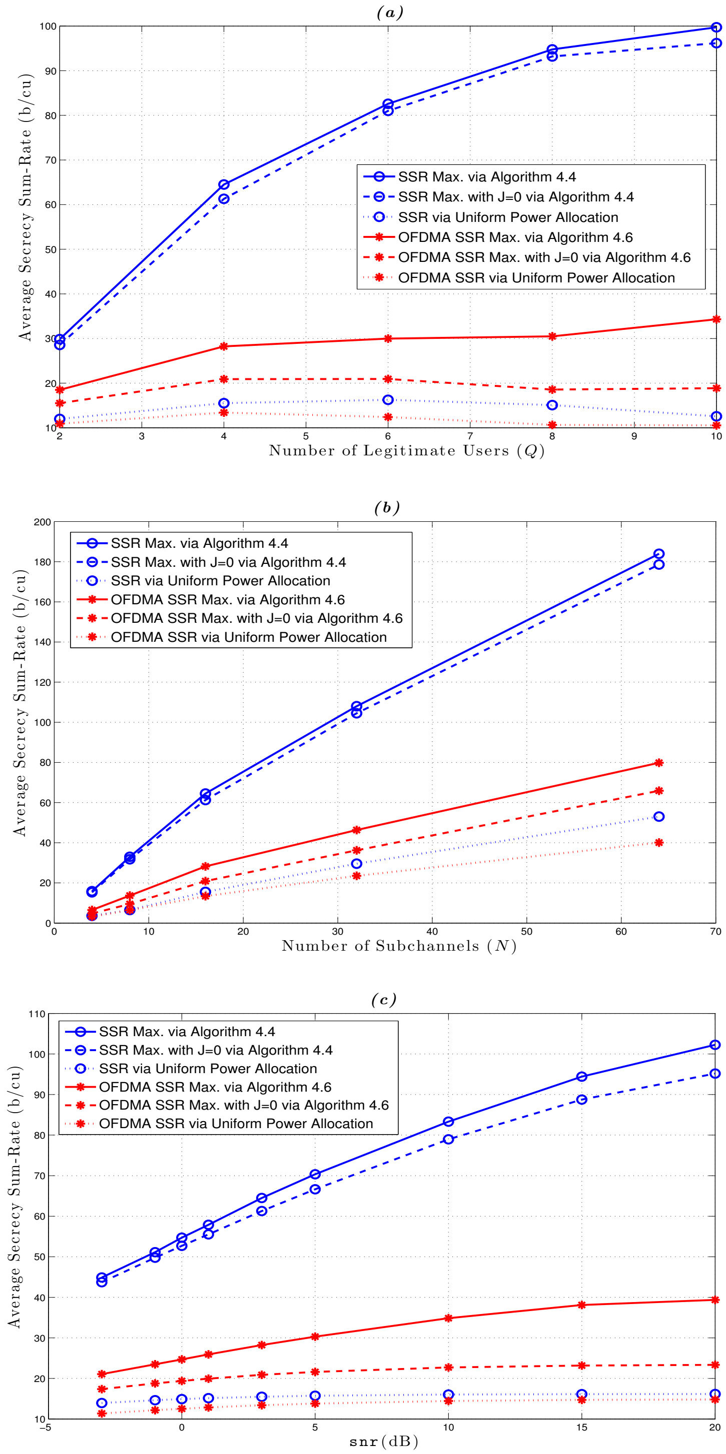

Figure 4.6: Average secrecy sum-rate (SSR) versus (a) number of legitimate users Q, (b) number of subchannels N, and (c) $\texttt{snr}$.

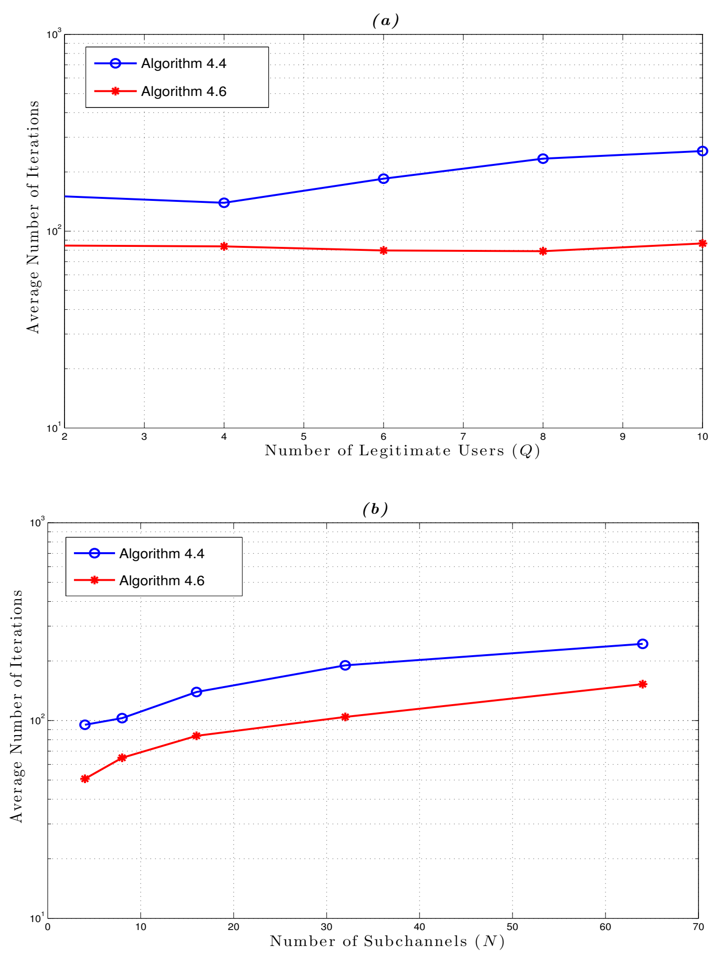

Figure 4.7: Average number of iterations for Algorithms 4.4 and 4.6 versus (a) number of legitimate users Q, and (b) number of subchannels N.

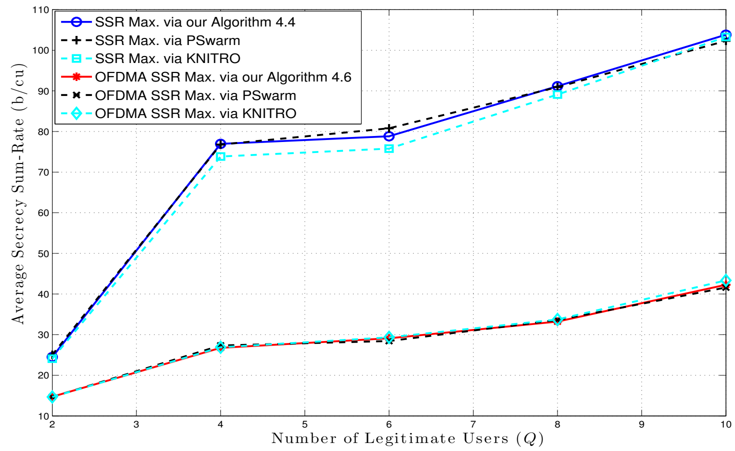

Figure 4.8: Comparison between the proposed algorithms versus other available schemes (adapted to our problem formulations) in terms of average secrecy sum-rate (SSR) achieved.

vergent to critical points of the DC-reformulation of (4.51) and (4.70), respectively. Hence, a good benchmark for our algorithms is to compare their achieved SSR with those attained by generic available methods attempting to compute locally or globally optimal solutions of such problems (but without rigorously verifying their optimality). In particular, we consider the algorithms KNITRO [16] and PSwarm [112] run over NEOS server [23]. It is worth mentioning that KNITRO is designed for smooth optimization, thus, in this case, we cast the SISO secrecy sum-rate maximization problem into the differentiable joint optimization reformulation introduced in Subsection 4.5.1.4. On the other side, PSwarm is suitable for both smooth and non-smooth optimization. In Figure 4.8 we report the average SSR achieved by the aforementioned algorithms (computed for 100 independent experiments) versus the number Q of main users ($J = Q/2$ and $N = 16$). This figure shows that our centralized and easily implementable Algorithm 4.4 outperforms KNITRO and has the same performance of the computationally very demanding PSwarm. Similarly, from Figure 4.8, it is also clear that Algorithm 4.6 achieves SSR that are comparable (and sometimes better) to those obtained from KNITRO and PSwarm; this means that, at least for this set of experiments, Algorithm 4.6 provides in a *distributed* way SSR that are very close to those obtained by *centralized* methods.

Finally, despite its simplicity, a case that is worth mentioning is that of zero friendly jammers and orthogonal subchannels. For this particular scenario, it

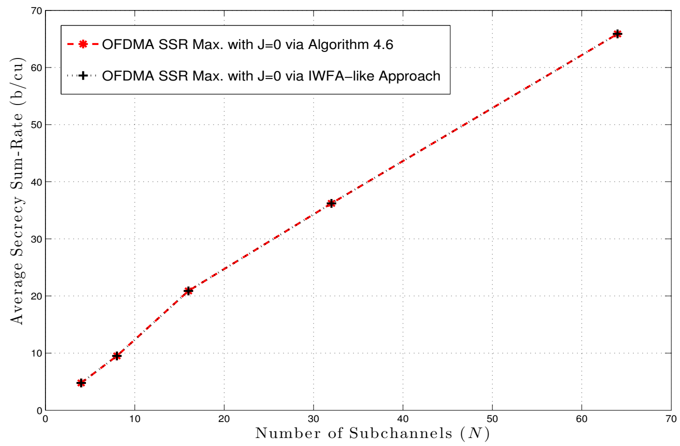

Figure 4.9: Comparison between Algorithm 4.6 and the approach proposed in [48] for the case of orthogonal subchannels and zero friendly jammers.

happens that we can adapt the results in [48] to our formulation in (4.70) and apply the iterative-waterfilling-like algorithm proposed in the cited reference to find an *optimal* power allocation for this optimization problem. Figure 4.9 shows that our Algorithm 4.6 (with $Q = 4$ main users) converges to such optimal power allocations, therefore it has the same performance (in terms of average secrecy sum-rates achieved) of the approach introduced in [48]. We recall that these averages were calculated over 50 independent channel realizations and 10 random channels assignments per channel realization.

4.5.5 The MIMO Case

We conclude the applications section showing that the spectrum management algorithms proposed for the SISO system implementing physical layer based security (see, Subsections 4.5.1 and 4.5.2) can be readily extended to the case where the network nodes are endowed with multiple antennas, that is, a MIMO communication system; see, e.g., [62, 79, 60].

Let us consider a wireless communication system composed of Q transmitter-receiver pairs (denoted by $q = 1, \ldots, Q$), J friendly jammers (denoted by $j = 1, \ldots, J$) and one eavesdropper (denoted by e). Different from the setting considered in the previous sections, we study a MIMO system where each legitimate transmitter, equipped with n_{T_q} antennas, aims to communicate a secrete message with its corresponding receiver (equipped with n_{R_q} antennas) in the presence of an eavesdropper that is also enabled with multiple n_{R_e} antennas. Furthermore, invoking the CJ paradigm, there are J friendly jammers equipped with $\widehat{n}_{T_j}$ antennas willing to help the legitimate users to secure their corresponding transmissions. In Figure 4.10, we illustrate this system for the simple case of a single legitimate pair, one friendly jammer and the eavesdropper. This figure also serves as an illustration of the channel gain matrices introduced below.

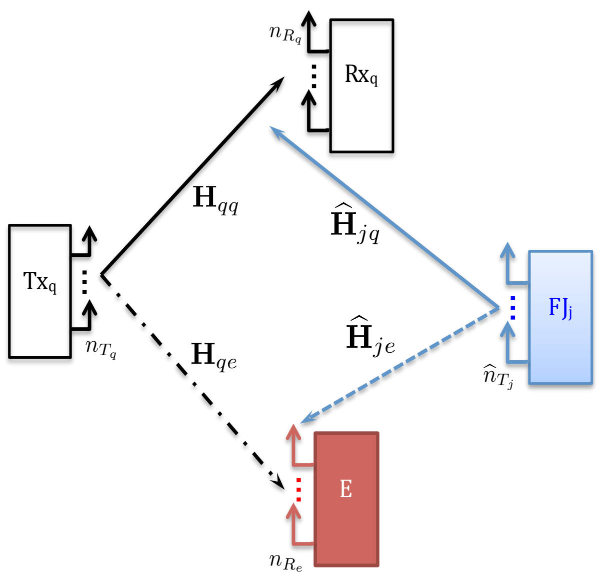

Figure 4.10: MIMO system composed of one legitimate pair, a single friendly jammer, and the eavesdropper. The arrows illustrate the channel matrices between the different entities in the network.

In the following discussion, we let:

- $\mathbf{Q}_q \in \mathbb{C}^{n_{T_q} \times n_{T_q}}$ denotes the transmit covariance matrix of the legitimate user q. It is worth emphasizing that, since $\mathbf{Q}_q$ is a covariance matrix then it is positive semidefinite i.e. $\mathbf{Q}_q \succeq \mathbf{0}$.
- $\widehat{\mathbf{Q}}_j \in \mathbb{C}^{\widehat{n}_{T_j} \times \widehat{n}_{T_j}}$ denotes the transmit covariance matrix of the friendly jammer j.
- $P_q^{\max}$ denotes the power budget of the q-th transmitter.
- $\widehat{P}_j^{\max}$ denotes the power budget of the j-th friendly jammer.
- $\mathbf{H}_{qq} \in \mathbb{C}^{n_{R_q} \times n_{T_q}}$ represents the channel matrix between the q-th legitimate transmitter and its intended receiver.

- $\mathbf{H}_{rq} \in \mathbb{C}^{n_{R_q} \times n_{T_r}}$ denotes the cross-channel matrix between the transmitter of the legitimate user r and the receiver of the legitimate user q. Likewise, $\mathbf{H}_{re} \in \mathbb{C}^{n_{R_e} \times n_{T_r}}$ denotes the cross-channel matrix between the transmitter of the legitimate user r and the receiver of the eavesdropper.

- $\widehat{\mathbf{H}}_{jq} \in \mathbb{C}^{n_{R_q} \times \widehat{n}_{T_j}}$ represents the cross-channel matrix between the transmitter of the friendly jammer j and the receiver of the legitimate user q. Similarly, $\widehat{\mathbf{H}}_{je} \in \mathbb{C}^{n_{R_e} \times \widehat{n}_{T_j}}$ denotes the cross-channel matrix between the transmitter of the friendly jammer j and the eavesdropper's receiver.

- $\mathbf{R}_{n_q} \in \mathbb{C}^{n_{R_q} \times n_{R_q}}$ represents the covariance matrix of the noise at the receiver of the legitimate user q, assumed to be positive definite.

As done in the previous section and in the related literature, we assume that CSI is available.

Using the notation introduced above, under basic information theoretical assumptions, the maximum achievable rate on link q for a given transmit covariance matrix profile $\left(\mathbf{Q} \triangleq (\mathbf{Q}_q)_{q=1}^{Q}, \widehat{\mathbf{Q}} \triangleq (\widehat{\mathbf{Q}}_j)_{j=1}^{J}\right)$ with each $\mathbf{Q}_q \succeq \mathbf{0}$ and $\widehat{\mathbf{Q}}_j \succeq \mathbf{0}$, is given by [22]

$$r_{qq}(\mathbf{Q}, \widehat{\mathbf{Q}}) = \log\det\left(\mathbf{I} + \mathbf{H}_{qq}^H \, \mathbf{R}_{-q}^{-1}(\mathbf{Q}_{-q}, \widehat{\mathbf{Q}}) \, \mathbf{H}_{qq} \, \mathbf{Q}_q\right),$$

where $\mathbf{Q}_{-q} \triangleq (\mathbf{Q}_r)_{r \neq q}$ denotes the covariance matrices of all users except the q-th one; and

$$\mathbf{R}_{-q}(\mathbf{Q}_{-q}, \widehat{\mathbf{Q}}) \triangleq \mathbf{R}_{n_q} + \sum_{r \neq q} \mathbf{H}_{rq} \mathbf{Q}_r \mathbf{H}_{rq}^H + \sum_{j=1}^{J} \widehat{\mathbf{H}}_{jq} \widehat{\mathbf{Q}}_j \widehat{\mathbf{H}}_{jq}^H.$$

Similarly, the rate on the link between the transmitter of the legitimate user q and the eavesdropper's receiver is

$$r_{qe}(\mathbf{Q}, \widehat{\mathbf{Q}}) = \log\det\left(\mathbf{I} + \mathbf{H}_{qe}^H \, \widehat{\mathbf{R}}_{-q}^{-1}(\mathbf{Q}_{-q}, \widehat{\mathbf{Q}}) \, \mathbf{H}_{qe} \, \mathbf{Q}_q\right),$$

where

$$\widehat{\mathbf{R}}_{-q}(\mathbf{Q}_{-q}, \widehat{\mathbf{Q}}) \triangleq \mathbf{R}_{n_e} + \sum_{r \neq q} \mathbf{H}_{re} \mathbf{Q}_r \mathbf{H}_{re}^H + \sum_{j=1}^{J} \widehat{\mathbf{H}}_{je} \widehat{\mathbf{Q}}_j \widehat{\mathbf{H}}_{je}^H.$$

Thus, the secrecy rate on link q is

$$r_q^s(\mathbf{Q}, \widehat{\mathbf{Q}}) \triangleq \left[r_{qq}(\mathbf{Q}, \widehat{\mathbf{Q}}) - r_{qe}(\mathbf{Q}, \widehat{\mathbf{Q}})\right]^+ \tag{4.82}$$

where $[\bullet]^+$ denotes the Euclidean projection onto $\mathbb{R}_+$.

Among the different system designs considered for the SISO case in the previous sections, let us focus here only on the secrecy sum-rate maximization problem for the MIMO system described above. Under the assumption that, each legitimate user q and every friendly jammer j have limited transmit power budgets i.e. $\text{tr}(\mathbf{Q}_q) \leq P_q^{\max}$ for all $q = 1, \ldots, Q$, and $\text{tr}(\mathbf{Q}_j) \leq \widehat{P}_j^{\max}$ for every $j = 1, \ldots, J$, the MIMO secrecy sum-rate maximization problem is given by

$$\begin{aligned} \underset{\mathbf{Q}, \widehat{\mathbf{Q}} \succeq \mathbf{0}}{\text{maximize}} \quad & r^s(\mathbf{Q}, \widehat{\mathbf{Q}}) \triangleq \sum_{q=1}^{Q} r_q^s(\mathbf{Q}, \widehat{\mathbf{Q}}) \\ \text{subject to} \quad & \text{tr}(\mathbf{Q}_q) \leq P_q^{\max} \quad \forall\, q = 1, \ldots, Q, \\ & \text{tr}(\mathbf{Q}_j) \leq \widehat{P}_j^{\max} \quad \forall\, j = 1, \ldots, J. \end{aligned} \tag{4.83}$$

By invoking the well-known Sylvester's Determinant Identity, it is not difficult to see that, for every $q = 1, \ldots, Q$, the secrecy rate on link q, r_q^s [cf. (4.82)] can be re-expressed as

$$r_q^s(\mathbf{Q}, \widehat{\mathbf{Q}}) \triangleq \left[f_{q,1}(\mathbf{Q}, \widehat{\mathbf{Q}}) + f_{q,2}(\mathbf{Q}, \widehat{\mathbf{Q}})\right]^+,$$

where

$$\begin{aligned} f_{q,1}(\mathbf{Q}, \widehat{\mathbf{Q}}) &\triangleq \log\det\left(\mathbf{R}_{-q}(\mathbf{Q}_{-q}, \widehat{\mathbf{Q}}) + \mathbf{H}_{qq}\mathbf{Q}_q\mathbf{H}_{qq}^H\right) + \log\det\left(\widehat{\mathbf{R}}_{-q}(\mathbf{Q}_{-q}, \widehat{\mathbf{Q}})\right) \\ f_{q,2}(\mathbf{Q}, \widehat{\mathbf{Q}}) &\triangleq -\log\det\left(\widehat{\mathbf{R}}_{-q}(\mathbf{Q}_{-q}, \widehat{\mathbf{Q}}) + \mathbf{H}_{qq}\mathbf{Q}_q\mathbf{H}_{qq}^H\right) - \log\det\left(\mathbf{R}_{-q}(\mathbf{Q}_{-q}, \widehat{\mathbf{Q}})\right). \end{aligned}$$

It is easy to show that each function $f_{q,1}$ is concave in $(\mathbf{Q}, \widehat{\mathbf{Q}})$, and similarly each $f_{q,2}$ is convex in $(\mathbf{Q}, \widehat{\mathbf{Q}})$, with each $\mathbf{Q}_q \succeq \mathbf{0}$ and $\widehat{\mathbf{Q}}_j \succeq \mathbf{0}$ in both cases. Hence, using the same argument as in the SISO secrecy sum-rate maximization problem (refer to Subsection 4.5.1) the problem in (4.83) can be rewritten equivalently as

$$\begin{array}{lll} \underset{\mathbf{Q},\widehat{\mathbf{Q}}\succeq\mathbf{0}}{\text{maximize}} & \displaystyle\sum_{q=1}^{Q} \underset{\lambda_q \in [0,1]}{\text{maximum}} \Big(\lambda_q \, f_{q,1}(\mathbf{Q},\widehat{\mathbf{Q}}) + \lambda_q \, f_{q,2}(\mathbf{Q},\widehat{\mathbf{Q}}) \Big) & \\ \text{subject to} & \operatorname{tr}(\mathbf{Q}_q) \leq P_q^{\max} \quad \forall\, q = 1,\ldots,Q, & \\ & \operatorname{tr}(\mathbf{Q}_j) \leq \widehat{P}_j^{\max} \quad \forall\, j = 1,\ldots,J, & \end{array} \tag{4.84}$$

which is clearly an instance of the MSM problem introduced in (4.2). As a direct consequence, since assumptions A1-A6 are satisfied by the problem above, we can then apply Algorithms 4.1 or 4.3 to compute critical points of the DC-reformulation of (4.84) or stationary points of its joint optimization reformulation, respectively. Thus, the results developed in the previous sections, for a SISO communication system, can be easily extended to allocate resources in the more complex case of a MIMO system implementing physical layer based security.

4.6 Conclusion

This chapter introduced a novel resource allocation problem in a multiuser system where the sum-utility function has the particular structure of the sum of continuous max functions, what we called the MSM problem. We proposed two different iterative algorithms with provable convergence to critical points (stationary points) of the DC-reformulation (JO-reformulation) of the MSM. Furthermore, those critical points (or stationary solutions) obtained from the proposed algorithms coincide under some conditions with d-stationary solutions of the MSM problem. The aforementioned theory is suitable to deal with resource allocation problems in the area of physical layer based security. Therefore, this chapter develops centralized and distributed algorithms for the secrecy sum-rate maximization problem (possibly with QoS constraints) and for the Max-Min fairness design, for a SISO wireless system composed of multiple legitimate users, multiple friendly jammers and a single eavesdropper, where the main users communicate over multiple subchannels. Further extensions include the MIMO system. To the best of our knowledge, these models are the most general ones analyzed so far in the related literature. Among the two approached proposed for addressing the aforementioned problems, our numerical experiments suggest that both schemes perform similarly in practice. It is worth remarking that, the DC-decomposition of the secrecy rate leads naturally to the design of algorithms for more complex models, such as those involving QoS constraints.

The design of *distributed* algorithms for the MSM problem, where the agents iteratively update their strategies either in parallel or sequentially is still an open question and we will continue to study it in our future work. We are also interested in continue to analyze the relations between the MSM and its reformulations. In particular, a more detailed study of the solution points obtained from the algorithms devised through those reformulations and their connection with the original MSM problem is also part of our future research.

Chapter 5

Conclusions

This dissertation provides several advances in the current literature of both areas of resource allocation problems and in that of signal processing in communication systems. However, most of the results presented in this work are quite general and can be easily extended to encompass a great range of resource allocation problems in other fields.

Chapter 2 presented an analysis based on LCP theory of the maximum sum-utility of a communication system (i.e. the sum-rate) achieved when a noncooperative approach is used to dynamically allocate the spectrum. In particular, we considered the case of unbounded power budgets as an approximation of systems endowed with large but finite power constraints. It is worth stressing that, different from other studies proposed in the literature, our results are not restricted in number of users and subchannels. The aforementioned model led us to derive an interesting conclusion regarding the efficiency of the NE in this context; namely, the system's utility obtained from the NE may be finite even when the users have infinite power budgets, as opposed to the infinite utility achieved when a centralized approach is used to allocate the power in the system. In simpler words, when the spectrum is managed through a game theoretical model, more power budget does not necessarily translates into larger transmission rates. This suggests that the selfish behavior of the players may not be overcome by the provision of infinite resources. Aside from the previous observations, the LCP framework developed in this chapter provided the tools to derive several sharper results. Among them, we highlight the following three. First, we were able to characterize the NE that yields an infinite sum-rate when the power budget is increased toward infinity. Second, we provided sufficient conditions under which only a finite number of NE exists. And third, we devised a case that prohibits the presence of the Braess-type paradox in this class of systems.

Chapter 3 introduced a class of provable convergent algorithms that are applicable to find stationary solutions of multiuser programs characterized by: differentiable, nonconvex and nonseparable objective function, and convex coupling constraints. Even though, we focussed our analysis to objective functions of the DC type, our results can be extended to treat sum-utility functions not necessarily in this form. Two remarkable features of our schemes are: first, they can be implemented in a fairly distributed way, thus, they are suitable for addressing resource allocation problems in large scale systems; and, second, they allow inexact computations, hence the computational effort can be reduced. Among the diverse applications of the proposed algorithms, we moved one step forward in the power allocation problem in a DSM framework by considering this problem in an OFDMA system implementing physical layer based security. This system consists of an arbitrary number of main users and friendly jammers, and a single eavesdropper; thus, extending the models considered so far in the literature. The system design was formulated as a game, which is of the generalized type. Different from previous works, we carefully addressed the nonconvexity and nondifferentiability of such a game via the introduction of relaxed equilibrium concepts. Interestingly, the proposed DC algorithms can be applied to find the aforementioned equilibrium points. We also considered the sum-rate maximization problem of MIMO CR systems, for which the algorithms developed in the literature lack of theoretical convergence. Nevertheless, our DC schemes can be applied to such a problem and thus, we have provided for the first time a class of distributed algorithms with provable convergence. It is worth mentioning that, numerically speaking, our experiments suggested that the proposed distributed algorithms have similar performance and sometimes better than centralized ones.

Finally, in Chapter 4 we studied a resource allocation problem in a multiuser system where the utility function is the sum of continuous max functions where the maximand has a particular structure. We called this maximization problem MSM. Toward attempting the solution of such a problem, we followed two approaches: (i) a DC-based, and (ii) a smooth joint optimization reformulation. These approaches led us to develop a family of SCA-based algorithms with provable convergence to critical points of the MSM problem's DC reformulation and to stationary solutions of the joint optimization reformulation of such a problem. We have also established some connections between the points obtained from the proposed schemes and the stationary solutions of the original problem. We applied our schemes to the power allocation problem in a communication system implementing physical layer based security, but different from the model considered in Chapter 3, the communication between the legitimate users is over multiple subchannels, for which the smooth reformulation introduced in the cited chapter is too restrictive. For the case of multiple orthogonal subchannels, the proposed algorithms are distributed, and for the more general case of multiple parallel subchannels the schemes are centralized. The nontrivial DC decomposition of the secrecy rate obtained in this chapter lead us also to devise iterative algorithms applicable to the well-known Max-Min fairness case and to the quality of service constrained problem. In terms of numerical results, the experiments suggested that (for the applications under consideration) the DC-based schemes and the joint optimization approach perform similarly both in terms of convergence speed and achieved secrecy sum-rate. The numerical experiments also indicate that our (distributed) easy-to-implement algorithms achieve objective values that are comparable to those obtained from (centralized) expensive approaches.

As a final overall remark, it is important to highlight that the introduction of optimization theory in different signal processing applications has motivated several advances in the former field. Figure 1.1 in Chapter 1 illustrates this observation by showing how the power allocation problem under different settings has created the necessity of introducing broader (but relaxed) equilibrium concepts in the area of Game Theory, ranging from the classical NE (in the context of convex games) to the concept of B-QNE (in the context of nonconvex and nondifferentiable games) introduced in this work; all of this with the objective of devising practical solutions for those applications. Consequently, it is clear that the research presented in this dissertation has provided significant advances in the aforementioned fields.

References

[1] H. Al-Shatri and T. Weber, *Achieving the maximum sum rate using D.C. programming in cellular networks*, IEEE Trans. Signal Process., 60 (2012), pp. 1331–1341.

[2] E. Altman, T. Jimenez, N. Vicuna, and R. Marquez, *Coordination games over collision channels*, in 6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks - WiOPT, Berlin, Germany, Apr. 2008, pp. 523–527.

[3] E. Altman, V. Kamble, and H. Kameda, *A braess type paradox in power control over interference channels*, in 6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks - WiOPT, Berlin, Germany, Apr. 2008, pp. 555–559.

[4] A. Alvarado, G. Scutari, and J.-S. Pang, *A new decomposition method for multiuser DC-programming and its applications*, IEEE Trans. Signal Process., 62 (2014), pp. 2984–2998.

[5] L. An and T. Pham Dinh, *The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems*, Annals of Operations Research, 133 (2005), pp. 23–46.

[6] E. Anshelevich, A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, and T. Roughgarden, *The price of stability for network design with fair cost allocation*, in 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS), Oct. 2004, pp. 59–73.

[7] J.-P. Aubin, *Mathematical methods of game and economic theory*, Courier Dover Publications, New York, NY, USA, Nov. 2007.

[8] O. Ben-Ayed and C. E. Blair, *Computational difficulties of bilevel linear programming*, Operations Research, 38 (1990), pp. 556–560.

[9] C. Berge, *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*, Oliver and Boyd, 1963.

[10] R. Berry and D. Tse, *Shannon meets nash on the interference channel*, IEEE Trans. Inf. Theory, 57 (2011), pp. 2821–2836.

[11] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex analysis and optimization*, Athena Scientific Belmont, 2003.

[12] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*, Athena Scientific Press, 2nd ed., 1989.

[13] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer Verlag, 2000.

[14] D. Braess, *Über ein paradoxon aus der verkehrsplanung*, Mathematical Methods of Operations Research, 12 (1968), pp. 258–268.

[15] G. Bresler and D. Tse, *The two-user gaussian interference channel: a deterministic view*, European transactions on telecommunications, 19 (2008), pp. 333–354.

[16] R. H. Byrd, J. Nocedal, and R. A. Waltz, *KNITRO: An integrated package for nonlinear optimization*, in Large-scale nonlinear optimization, Springer, 2006, pp. 35–59.

[17] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen, *Autonomous spectrum balancing for digital subscriber lines*, IEEE Trans. Signal Process., 55 (2007), pp. 4241–4257.

[18] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, *Optimal multiuser spectrum management for digital subscriber lines*, IEEE Trans. Commun., 54 (2006), pp. 922–933.

[19] V. Chan and W. Yu, *Joint multiuser detection and optimal spectrum balancing for digital subscriber lines*, EURASIP journal on applied signal processing, 2006 (2006), pp. 1–13.

[20] S. Chung, S. Kim, J. Lee, and J. Cioffi, *A game-theoretic approach to power allocation in frequency-selective gaussian interference channels*, in Proceedings of IEEE International Symposium on Information Theory, Jun. 2003, pp. 316–316.

[21] R. Cottle, J.-S. Pang, and R. Stone, *The linear complementarity problem*, no. 60, Society for Industrial and Applied Mathematics, 2009.

[22] T. Cover and J. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[23] J. Czyzyk, M. P. Mesnier, and J. J. Moré, *The NEOS server*, IEEE Comput. Sci. Eng., 5 (1998), pp. 68–75.

[24] L. Dong, Z. Han, A. Petropulu, and H. Poor, *Improving wireless physical layer security via cooperating relays*, IEEE Trans. Signal Process., 58 (2010), pp. 1875–1888.

[25] R. Etkin, A. Parekh, and D. Tse, *Spectrum sharing for unlicensed bands*, IEEE J. Sel. Areas Commun., 25 (2007), pp. 517–528.

[26] F. Facchinei, A. Fischer, and V. Piccialli, *On generalized nash games and variational inequalities*, Operations Research Letters, 35 (2007), pp. 159–164.

[27] F. Facchinei and C. Kanzow, *Generalized nash equilibrium problems*, A Quart. J. Operat. Res. (4OR), 5 (2007), pp. 173–210.

[28] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problem*, Springer-Verlag, New York, 2003.

[29] F. Facchinei and J.-S. Pang, *Nash equilibria: the variational approach*, in Convex optimization in signal processing and communications, London, 2009, ch. 12, Cambridge Univ. Press, pp. 443–493.

[30] F. Facchinei, J.-S. Pang, G. Scutari, and L. Lampariello, *VI-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks*, Mathematical Programming, (2013), pp. 1–38.

[31] F. Facchinei, V. Piccialli, and M. Sciandrone, *Decomposition algorithms for generalized potential games*, Computational Optimization and Applications, 50 (2011), pp. 237–262.

[32] D. H. Fang and X. P. Zhao, *Local and global optimality conditions for DC infinite optimization problems*, Taiwanese Journal of Mathematics, (2013. Available at: http://journal.taiwanmathsoc.org.tw.).

[33] R. H. Gohary, Y. Huang, Z.-Q. Luo, and J.-S. Pang, *A generalized iterative water-filling algorithm for distributed power control in the presence of a jammer*, IEEE Transactions on Signal Processing, 57 (2009), pp. 2660–2674.

[34] O. Güler, *Foundations of optimization*, vol. 258, Springer, 2010.

[35] Z. Han, N. Marina, M. Debbah, and A. Hjørungnes, *Improved wireless secrecy rate using distributed auction theory*, in 5th International Conference on Mobile Ad-hoc and Sensor Networks, IEEE, 2009, pp. 442–447.

[36] Z. Han, N. Marina, M. Debbah, and A. Hjørungnes, *Physical layer security game: interaction between source, eavesdropper, and friendly jammer*, EURASIP J. Wireless Commun. Netw., 2009 (2009), pp. 11:1–11:10.

[37] S. Hayashi and Z.-Q. Luo, *Spectrum management for interference-limited multiuser communication systems*, IEEE Trans. Inf. Theory, 55 (2009), pp. 1153–1175.

[38] S. Haykin, *Cognitive radio: brain-empowered wireless communications*, IEEE Journal on Selected Areas in Communications, 23 (2005), pp. 201–220.

[39] X. He and A. Yener, *Cooperative jamming: The tale of friendly interference for secrecy*, in Securing Wireless Communications at the Physical Layer, Springer, 2010, pp. 65–88.

[40] J.-B. Hiriart-Urruty, *Generalized differentiability, duality and optimization for problems dealing with differences of convex functions*, in Convexity and duality in optimization, Springer, 1985, pp. 37–70.

[41] M. Hong and Z.-Q. Luo, *Signal processing and optimal resource allocation for the interference channel*, Elsevier e-Reference-Signal Processing, (2013. [Online]. Available: http://arxiv.org/pdf/1206.5144v1.pdf.).

[42] R. Horst, P. Pardalos, and N. Thoai, *Introduction to global optimization*, vol. 48, Kluwer Academic Publishers, 2nd ed., 2000.

[43] R. Horst and N. V. Thoai, *DC programming: overview*, J. Optim. Theory Appl., 103 (1999), pp. 1–43.

[44] J. Huang, R. A. Berry, and M. L. Honig, *Distributed interference compensation for wireless networks*, IEEE Jour. on Selected Areas in Communications, 24 (2006), pp. 1074–1084.

[45] S. Huberman, C. Leung, and T. Le-Ngoc, *Dynamic spectrum management (dsm) algorithms for multi-user xdsl*, IEEE Communications Surveys & Tutorials, 14 (2012), pp. 109–130.

[46] S. Jafar and M. Fakhereddin, *Degrees of freedom for the MIMO interference channel*, IEEE Trans. Inf. Theory, 53 (2007), pp. 2637–2642.

[47] E. Jorswieck, A. Wolf, and S. Gerbracht, *Secrecy on the Physical Layer in Wireless Networks*, INTECH, 2010, ch. 20, pp. 413–435.

[48] E. A. Jorswieck and A. Wolf, *Resource allocation for the wire-tap multi-carrier broadcast channel*, in International Conference on Telecommunications (ICT) 2008, IEEE, 2008, pp. 1–6.

[49] F. P. Kelly, A. K. Maulloo, and D. K. Tan, *Rate control for communication networks: shadow prices, proportional fairness and stability*, Journal of the Operational Research society, (1998), pp. 237–252.

[50] A. Khabbazibasmenj, F. Roemer, S. Vorobyov, and M. Haardt, *Sum-rate maximization in two-way AF MIMO relaying: Polynomial time solutions to a class of DC programming problems*, IEEE Trans. Signal Process., 60 (2012), pp. 5478–5493.

[51] S.-J. Kim and G. B. Giannakis, *Optimal resource allocation for MIMO ad hoc cognitive radio networks*, IEEE Trans. Inf. Theory, 57 (2011), pp. 3117–3131.

[52] E. Koutsoupias and C. Papadimitriou, *Worst-case equilibria*, in Proceedings of the 16th annual conference on Theoretical aspects of computer science, Springer-Verlag, Mar. 1999, pp. 404–413.

[53] G. R. Lanckriet and B. K. Sriperumbudur, *On the convergence of the concave-convex procedure*, in Advances in neural information processing systems, 2009, pp. 1759–1767.

[54] E. Larsson, E. Jorswieck, J. Lindblom, and R. Mochaourab, *Game theory and the flat-fading gaussian interference channel*, IEEE Signal Process. Mag., 26 (2009), pp. 18–27.

[55] H. Le Thi, T. Pham Dinh, and M. Le Dung, *Exact penalty in DC programming*, Vietnam Journal of Mathematics, 27 (1999), pp. 169–178.

[56] H. A. Le Thi and T. P. Dinh, *DC programming in communication systems: challenging problems and methods*, Vietnam Journal of Computer Science, (2013), pp. 1–14.

[57] H. A. Le Thi, T. P. Dinh, and H. Van Ngai, *Exact penalty and error bounds in DC programming*, Journal of Global Optimization, 52 (2012), pp. 509–535.

[58] A. Leshem and E. Zehavi, *Game theory and the frequency selective interference channel*, IEEE Signal Process. Mag., 26 (2009), pp. 28–40.

[59] J. Li, A. Petropulu, and S. Weber, *On cooperative relaying schemes for wireless physical layer security*, IEEE Trans. Signal Process., 59 (2011), pp. 4985–4997.

[60] Q. Li, M. Hong, H.-T. Wai, Y.-F. Liu, W.-K. Ma, and Z.-Q. Luo, *Transmit solutions for MIMO wiretap channels using alternating optimization*, IEEE J. on Selec. Areas in Comm., 31 (2013), pp. 1714–1727.

[61] Y. Liang, H. V. Poor, and S. Shamai (Shitz), *Information theoretic security*, Found. Trends Commun. Inf. Theory, 5 (2009), pp. 355–580.

[62] J. Liu, Y. T. Hou, and H. D. Sherali, *Optimal power allocation for achieving perfect secrecy capacity in MIMO wire-tap channels*, in 43rd Annual Conference on Information Sciences and Systems (CISS), IEEE, Mar. 2009, pp. 606–611.

[63] R. Liu and W. Trappe, *Securing Wireless Communications at the Physical Layer*, Incorporated Springer Publishing Company, 1st ed., 2009.

[64] A. Lozano, A. Tulino, and S. Verdu, *High-SNR power offset in multiantenna communication*, IEEE Trans. Inf. Theory, 51 (2005), pp. 4134–4151.

[65] R. Lui and W. Yu, *Low-complexity near-optimal spectrum balancing for digital subscriber lines*, in IEEE International Conference on Communications (ICC), vol. 3, May 2005, pp. 1947–1951.

[66] Z. Luo, J.-S. Pang, and D. Ralph, *Mathematical Programs With Equilibrium Constraints*, Cambridge University Press, 1996.

[67] Z.-Q. Luo, *Analysis of iterative waterfilling algorithm for multi-user spectrum management in digital subscriber lines.* private communication, May 2005.

[68] Z.-Q. Luo and J.-S. Pang, *Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines*, EURASIP Journal on Applied Signal Processing, 2006 (2006), pp. 1–10.

[69] Z.-Q. Luo and S. Zhang, *Dynamic spectrum management: Complexity and duality*, IEEE Journal of Selected Topics in Signal Processing,, 2 (2008), pp. 57–73.

[70] P.-E. Maingé and A. Moudafi, *Convergence of new inertial proximal methods for DC programming*, SIAM Journal on Optimization, 19 (2008), pp. 397–413.

[71] I. Malanchini, S. Weber, and M. Cesana, *Nash equilibria for spectrum sharing of two bands among two players*, in 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Sept. 2010, pp. 783–790.

[72] B. R. Marks and G. P. Wright, *A general inner approximation algorithm for nonconvex mathematical programs*, Operations Research, 26 (1978), pp. 681–683.

[73] R. Mochaourab and E. Jorswieck, *Resource allocation in protected and shared bands: uniqueness and efficiency of nash equilibria*, in Proceedings of the 4th. International ICST Conference on Performance Evaluation Methodologies and Tools, VALUETOOLS '09, Brussels, Belgium, Oct. 2009, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 68:1–68:10.

[74] A. Moudafi, *On critical points of the difference of two maximal monotone operators*, Afrika Matematika, (2013), pp. 1–7.

[75] A. Moudafi and P. Mainge, *On the convergence of an approximate proximal method for DC functions*, Journal of computational Mathematics, 24 (2006), pp. 475–480.

[76] J. Nash, *Equilibrium points in n-person games*, Proceedings of the national academy of sciences, 36 (1950), pp. 48–49.

[77] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course (Applied Optimization)*, Springer, 2004.

[78] L. J. Neumann and O. Morgenstern, *Theory of games and economic behavior*, vol. 60, Princeton University Press, Princeton, NJ, 1944.

[79] F. Oggier and B. Hassibi, *The secrecy capacity of the MIMO wiretap channel*, IEEE Trans. Inf. Theory, 57 (2011), pp. 4961–4972.

[80] M. J. Osborne and A. Rubinstein, *A course in game theory*, MIT press, Cambridge, MA, Jul. 2004.

[81] D. Palomar and M. Chiang, *Alternative distributed algorithms for network utility maximization: Framework and applications*, IEEE Trans. Autom. Control, 52 (2007), pp. 2254–2269.

[82] D. P. Palomar and M. Chiang, *A tutorial on decomposition methods for network utility maximization*, IEEE Journal on Selected Areas in Communications, 24 (2006), pp. 1439–1451.

[83] D. P. Palomar and M. Chiang, *Alternative distributed algorithms for network utility maximization: Framework and applications*, IEEE Transactions on Automatic Control, 52 (2007), pp. 2254–2269.

[84] J. Pang and G. Scutari, *Nonconvex games with side constraints*, SIAM J. on Optimization, 21 (2011), pp. 1491–1522.

[85] J.-S. Pang and G. Scutari, *Joint sensing and power allocation in nonconvex cognitive radio games: Quasi-nash equilibria*, IEEE Trans. Signal Process., 61 (2013), pp. 2366–2382.

[86] J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang, *Distributed power allocation with rate constraints in gaussian parallel interference channels*, IEEE Transactions on Information Theory, 54 (2008), pp. 3471–3489.

[87] J.-S. Pang, G. Scutari, D. Palomar, and F. Facchinei, *Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach*, IEEE Transactions on Signal Processing, 58 (2010), pp. 3251–3271.

[88] K. Phan, S. Vorobyov, C. Telambura, and T. Le-Ngoc, *Power control for wireless cellular systems via D.C. programming*, in IEEE/SP 14th Workshop on Statistical Signal Processing, 2007, pp. 507–511.

[89] T. D. Quoc and M. Diehl, *Sequential convex programming methods for solving nonlinear optimization problems with DC constraints.* Available at: http://arxiv.org/abs/1107.5841v1, 2011.

[90] M. Razaviyayn, M. Hong, and Z.-Q. Luo, *A unified convergence analysis of block successive minimization methods for nonsmooth optimization*, Arxiv.org, (2012).

[91] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[92] L. Rose, S. Perlaza, and M. Debbah, *On the nash equilibria in decentralized parallel interference channels*, in IEEE International Conference on Communications Workshops (ICC), Jun. 2011, pp. 1–6.

[93] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, *Distributed resource allocation schemes: Pricing algorithms for power control and beamformer design in interference networks*, IEEE Signal Processing Magazine, 26 (2009), pp. 53–63.

[94] G. Scutari, F. Facchinei, P. Song, D. Palomar, and J. Pang, *Decomposition by partial linearization: Parallel optimization of multi-agent systems*, IEEE Trans. Signal Process., 62 (2014), pp. 641–656.

[95] G. Scutari, D. Palomar, and S. Barbarossa, *Asynchronous iterative water-filling for gaussian frequency-selective interference channels*, IEEE Trans. Inf. Theory, 54 (2008), pp. 2868–2878.

[96] ——, *Optimal linear precoding strategies for wideband non-cooperative systems based on game theory - part II: Algorithms*, IEEE Trans. Signal Process., 56 (2008), pp. 1250–1267.

[97] ——, *Optimal linear precoding strategies for wideband noncooperative systems based on game theory - part I: Nash equilibria*, IEEE Trans. Signal Process., 56 (2008), pp. 1230–1249.

[98] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, *Convex optimization, game theory, and variational inequality theory in multiuser communication systems*, IEEE Signal Process. Mag., 27 (2010), pp. 35–49.

[99] G. Scutari, D. P. Palomar, J.-S. Pang, and F. Facchinei, *Flexible design of cognitive radio wireless systems*, IEEE Signal Process. Mag., 26 (2009), pp. 107–123.

[100] G. Scutari and J.-S. Pang, *Joint sensing and power allocation in nonconvex cognitive radio games: Nash equilibria and distributed algorithms*, IEEE Trans. Inf. Theory, (2013).

[101] A. J. Smola, S. Vishwanathan, and T. Hoffman, *Kernel methods for missing variables*, In Proc. of the 10th. International Workshop on Artificial Intelligence and Statistics, AISTATS05, R. Cowell and Z. Ghahramani (Eds.) (2005), pp. 325–332.

[102] K. Song, S. Chung, G. Ginis, and J. Cioffi, *Dynamic spectrum management for next-generation DSL systems*, IEEE Commun. Mag., 40 (2002), pp. 101–109.

[103] B. K. Sriperumbudur and G. R. Lanckriet, *A proof of convergence of the concave-convex procedure using Zangwill's theory*, Neural computation, 24 (2012), pp. 1391–1407.

[104] I. Stanojev and A. Yener, *Improving secrecy rate via spectrum leasing for friendly jamming*, IEEE Trans. Wireless Commun., 12 (2013), pp. 134–145.

[105] M. Su and H. K. Xu, *Remarks on the gradient-projection algorithm*, Journal of Nonlinear Analysis and Optimization: Theory & Applications, 1 (2010), pp. 35–43.

[106] W.-y. Sun, R. J. Sampaio, and M. Candido, *Proximal point algorithm for minimization of DC function*, Journal of computational Mathematics, 21 (2003), pp. 451–462.

[107] C. W. Tan, D. P. Palomar, and M. Chiang, *Distributed optimization of coupled systems with applications to network utility maximization*, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, France, May 2006, IEEE, pp. 981–984.

[108] P. D. Tao and L. T. H. An, *Convex analysis approach to DC programming: theory, algorithms and applications*, Acta Mathematica Vietnamica, 22 (1997), pp. 289–355.

[109] ——, *A DC optimization algorithm for solving the trust-region subproblem*, SIAM Journal on Optimization, 8 (1998), pp. 476–505.

[110] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge university press, Cambridget, U.K., 2005.

[111] P. Tsiaflakis, M. Diehl, and M. Moonen, *Distributed spectrum management algorithms for multiuser DSL networks*, IEEE Trans. Signal Process., 56 (2008), pp. 4825–4843.

[112] A. I. F. Vaz and L. Vicente, *Pswarm: A hybrid solver for linearly constrained global derivative-free optimization*, Optimization Methods & Software, 24 (2009), pp. 669–685.

[113] N. Vucic, S. Shi, and M. Schubert, *DC programming approach for resource allocation in wireless networks*, in 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010, pp. 380–386.

[114] X. Wang, M. Tao, J. Mo, and Y. Xu, *Power and subcarrier allocation for physical-layer security in OFDMA-ased broadband wireless networks*, IEEE Transactions on Information Forensics and Security, 6 (2011), pp. 693–702.

[115] Y. Wu and K. J. R. Liu, *An information secrecy game in cognitive radio networks*, IEEE Trans. Inf. Forensics Security, 6 (2011), pp. 831–842.

[116] A. Wyner, *The wire-tap channel*, Bell System Technical Journal, 54 (1975), pp. 1355–1387.

[117] Y. Xu, T. Le-Ngoc, and S. Panigrahi, *Global concave minimization for optimal spectrum balancing in multi-user DSL networks*, IEEE Trans. Signal Process., 56 (2008), pp. 2875–2885.

[118] N. Yamashita and Z.-Q. Luo, *A nonlinear complementarity approach to multiuser power control for digital subscriber lines*, Optimization Methods and Software, 19 (2004), pp. 633–652.

[119] Y. Yang, Q. Li, W.-K. Ma, J. Ge, and P. Ching, *Cooperative secure beamforming for AF relay networks with multiple eavesdroppers*, IEEE Signal Process. Lett., 20 (2013), pp. 35–37.

[120] W. Yu, G. Ginis, and J. Cioffi, *Distributed multiuser power control for digital subscriber lines*, IEEE J. Sel. Areas Commun., 20 (2002), pp. 1105–1115.

[121] W. Yu and R. Lui, *Dual methods for nonconvex spectrum optimization of multicarrier systems*, IEEE Trans. Commun., 54 (2006), pp. 1310–1322.

[122] A. L. Yuille and A. Rangarajan, *The concave-convex procedure*, Neural Computation, 15 (2003), pp. 915–936.

[123] W. I. Zangwill, *Nonlinear programming: a unified approach*, Prentice-Hall Englewood Cliffs, NJ, 1969.

[124] R. Zhang, Y.-C. Liang, and S. Cui, *Dynamic resource allocation in cognitive radio networks*, IEEE Signal Processing Magazine, 27 (2010), pp. 102–114.

[125] R. Zhang, L. Song, Z. Han, and B. Jiao, *Improve physical layer security in cooperative wireless network using distributed auction games*, in IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)., IEEE, 2011, pp. 18–23.

[126] Y. Zhang, E. Dall'Anese, and G. B. Giannakis, *Distributed optimal beamformers for cognitive radios robust to channel uncertainties*, IEEE Trans. Signal Process., 60 (2012), pp. 6495–6508.

[127] Q. Zhao and B. M. Sadler, *A survey of dynamic spectrum access*, IEEE Signal Process. Mag., 24 (2007), pp. 79–89.